HUMAN BIOLOGY
a record of research
FEBRUARY, 1949
VOL. 21 No. 1

A PRACTICAL METHOD OF ESTIMATING THE MEAN AND STANDARD DEVIATION OF TRUNCATED NORMAL DISTRIBUTIONS*

BY JOHANNES IPSEN, JR.

MANY distributions of biological data present themselves as truncated, i. e., some of the observations in a sample are either below or above a limit which is inherent in the method of observation. The values beyond this limit are believed to form a continuation of the scale of measurement which is within the reach of the method of measurement.

The problem of estimating the standard deviation and mean of the population from which the sample is taken is somewhat different from the problem solved by Pearson and Lee in dealing with incomplete distributions where the method of sampling curtails the distribution at a certain point. In their case the proportion in the missing tail must also be estimated from the known tail.

The present case is exemplified by the frequent event in experimental biology, that a known number, n, of individuals are exposed to some agent, and the responses of some of these, a, fall outside the truncation point. For instance: 20 animals are injected with the same dose of antigen, blood samples from each animal are tested for antibody after a certain time, but only 12 of these samples contain measurable amounts,

* From the Department of Biostatistics, Harvard School of Public Health.

while 8 samples do not display any antibody with the given technic. It is reasonable to believe that these 8 animals have also responded to the antigen, but that the amount of antibody produced is too small to be demonstrated by the given method.

If the mean response of these 20 animals should be compared to the mean of another group treated with a more potent antigen which is able to produce measurable antibodies in all individuals, it is important to estimate the mean and standard deviation in the truncated sample by a method for which the error of the estimates is known.

Pearson and Lee could only utilize the information given by the recorded observation, the size of the whole group being unknown. In our case we can operate with both the measurements of the 12 individuals and the fact that 8 out of 20 animals have responded less than the point of truncation.

It is suggested that the data be arranged in the following form: (1) The scale of measurement is translated so that the point of truncation is zero. The measurements are then given in terms of

$$r_i = u_i - r_0$$

where the u_i are the actual measurements or their transformation into normally distributed metameters, and r_0 is the point of truncation. r_0 will always be chosen on the basis of the biological theory or the praxis underlying the type of experiment; e. g., if an antibody cannot be measured in titers less than 2, the natural point of truncation is 2. (2) All observations in the truncated part are given the value 0, although it is believed that the theoretical values are less than 0.

It will be convenient to consider first some of the mathematical relations which hold in the parent population. We assume first a normal distribution with mean μ and unit variance, truncated at 0. We recall that the value 0 is assigned to all negative values of r.

Then the fraction P below the truncation point is

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-(r-\mu)^2/2} dr = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\mu} e^{-t^2/2} dt. \qquad (1)$$

The truncated mean, ρ, is determined by

$$\rho = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} r e^{-(r-\mu)^2/2} dr = \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2} + \mu Q, \qquad (2)$$

and introducing

$$z = \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2} \tag{3}$$

(2) becomes

$$\rho = z + \mu Q. \tag{4}$$

A convenient parameter will be

$$\gamma = \rho - \mu \tag{5}$$

for the difference between the truncated mean and the true mean. (Note that the sign of γ is positive if the truncation is at the lower end of the distribution, negative if the truncation is at the upper end of the distribution.)

Relations (1) to (5) serve to express any of the variables P, μ, ρ, and γ in terms of any other. In Table 1, ρ is taken as the independent variable.

Consider now the truncated variance, η, defined as

$$\eta = \int_{-\infty}^{\infty} (r-\rho)^2 dF(r) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} r^2 e^{-(r-\mu)^2/2} dr - \rho^2 \tag{6}$$

It is easy to show that

$$\eta = Q - \rho^2 + \mu\rho$$

or

$$\eta = Q - \rho\gamma. \tag{7}$$

If now we consider a parent population which is normal with mean $\bar{m}$ and variance σ^2, we have the following relations:

$$\begin{aligned} P' &= P, \\ \rho' &= \rho\sigma, \\ \eta' &= \eta\sigma^2, \end{aligned} \tag{8}$$

where the primed quantities refer to the distribution $(\bar{m}, \sigma^2)$, and the unprimed to the distribution $(\mu = \bar{m}/\sigma, 1)$.

Further we have obviously

$$\eta'/\rho'^2 = \eta/\rho^2 = \kappa, \tag{9}$$

say, a relation which will be useful later. It is clear that η and κ may be expressed as functions of P or of the other variables. In Table 2, κ is taken as the independent variable.

TABLE 1

Constants of a truncated normal distribution; standard deviation known from outside sources

ρ	μ	P	γ	$W(\mu)$
3.0	2.99962	.00135	.000382	.9998
2.5	2.49817	.00624	.00183	.9988
2.0	1.99131	.02322	.00869	.9943
1.9	1.88862	.02947	.01138	.9921
1.8	1.78518	.03712	.01482	.9902
1.7	1.68084	.04640	.01916	.9873
1.6	1.57574	.05758	.02426	.9823
1.5	1.46852	.07098	.03148	.9787
1.4	1.35998	.08692	.04002	.9727
1.3	1.24934	.10577	.05066	.9652
1.2	1.13614	.12795	.06386	.9560
1.1	1.01976	.15392	.08024	.9446
1.0	.89947	.1842	.10053	.9303
.9	.77423	.2194	.12577	.9121
.8	.64275	.2602	.15725	.8913
.7	.50319	.3074	.19681	.8645
.6	.35293	.3621	.24707	.8310
.50	.18810	.4254	.31190	.7886
.45	.09826	.4609	.35174	.7631
.40	.00213	.4992	.39787	.7340
.35	—.10204	.5406	.45204	.7007
.30	—.21651	.5857	.51651	.6618
.25	—.34486	.6349	.59486	.6160
.20	—.49288	.6890	.69288	.5609
.18	—.55969	.7122	.73969	.5355
.16	—.63225	.7364	.79225	.5077
.14	—.71198	.7618	.85198	.4771
.12	—.8010	.7884	.9210	.4429
.10	—.9023	.8166	1.0023	.4044
.09	—.9592	.8313	1.0492	.3831
.08	—1.0213	.8464	1.1013	.3602
.07	—1.0899	.8621	1.1599	.3353
.06	—1.1671	.8784	1.2271	.3085
.05	—1.2556	.8954	1.3056	.2782
.04	—1.3602	.9131	1.4002	.2444
.03	—1.4897	.93185	1.5197	.2058
.02	—1.6631	.95185	1.6831	.1600
.01	—1.9384	.97371	1.9484	.1001
.005	—2.1920	.98581	2.1970	.0623
.001	—2.7178	.99671	2.7188	.0183

TABLE 2

Constants of a truncated normal distribution; standard deviation estimated from truncated distribution

κ, h	P	ρ	γ	η	$W(\mu)$	$W(\sigma)$
.10	.00079	3.1585	.000216	.9985	.9999	1.8181
.15	.00507	2.5743	.001606	.9911	.9988	1.7391
.20	.01378	2.2085	.004836	.9755	.9970	1.6667
.30	.04050	1.7613	.01638	.9307	.9891	1.5384
.40	.07380	1.4810	.03295	.8774	.9770	1.4286
.50	.10928	1.2830	.05270	.8231	.9620	1.3333
.60	.14474	1.1336	.07433	.7710	.9448	1.2500
.70	.17909	1.0158	.09701	.7224	.9260	1.1764
.80	.21182	.9203	.12018	.6776	.9060	1.1111
.90	.24272	.8410	.14347	.6366	.8852	1.0526
1.00	.27177	.7741	.16663	.5992	.8638	1.0000
1.2	.32451	.6671	.21200	.5341	.8198	.9091
1.4	.37080	.5853	.25557	.4796	.7753	.8333
1.6	.41152	.5207	.29706	.4338	.7313	.7692
1.8	.44750	.4684	.33642	.3949	.6886	.7143
2.0	.47947	.4252	.37373	.3616	.6476	.6667
2.1417	.50000	.39894	.39894	.34085	.6198	.6366
2.5	.54549	.3444	.45869	.2965	.5544	.5714
3.0	.59679	.2883	.53338	.2494	.4755	.5000
4.0	.67109	.2160	.65892	.1866	.3558	.4000
5.0	.72223	.1716	.76106	.1472	.2743	.3333
6.0	.75955	.1417	.84657	.1205	.2178	.2857
7.0	.78799	.1203	.91978	.1013	.1773	.2500
8.0	.81037	.1043	.98357	.0870	.1476	.2222
9.0	.82847	.0919	1.04002	.0760	.1251	.2000
10	.84336	.0820	1.0904	.06723	.1078	.1818
12	.86652	.0673	1.1774	.05429	.0829	.1538
14	.88368	.0568	1.2504	.04524	.0664	.1333
16	.89690	.0491	1.3132	.03860	.0547	.1176
18	.90747	.0431	1.3685	.03351	.0460	.1053
20	.91596	.0384	1.4168	.02957	.0396	.0952
30	.94244	.0246	1.6002	.01817	.0221	.0645
40	.95618	.0179	1.7260	.01286	.0148	.0488
60	.97030	.0115	1.8968	.00791	.0084	.0328
80	.97752	.0084	2.0134	.00561	.0057	.0247
100	.98191	.0066	2.1014	.00431	.0042	.0198
500	.99628	.0011	2.6779	.00065	.0005	.0040

Let the experiment provide a sample of n observations, of which a fall in the truncated portion, and are therefore scored 0. The procedure involves the computation of the sample mean and variance,

$$\bar{r} = \Sigma(r)/n$$

and

$$s^2 = [\Sigma(r^2) - n\bar{r}^2]/(n-1)$$

and of the ratio

$$h = s^2/\bar{r}^2.$$

The corresponding quantities from sample and population are then:

Quantity	*Sample*	*Population*
Fraction below $r = 0$	a/n	P
Mean	$\bar{r}$	$\rho' = \rho\sigma$
Variance	s^2	$\eta' = \eta\sigma^2$
Variance/(mean)^2	$h = s^2/\bar{r}^2$	$\kappa = \eta/\rho^2$

The procedure of estimating the population parameters will then be to equate appropriate sample values to the corresponding population values. There are two cases to consider, depending on whether the population standard deviation is known or unknown.

THE STANDARD DEVIATION IS KNOWN

This situation is not uncommon in the field of bioassays. In cases of rectilinear reaction curves the standard deviation of the observations around the line is often found to be independent of the mean reaction throughout the range of measurement, and also so at the end of the reaction line where it approaches some limit due to the technic of measurement. Some of the groups in this end may be truncated, and the standard deviation can then be computed from the complete distributions and later applied to the truncated ones. It should be emphasized that the standard deviation should also be estimated from the truncated distributions themselves and compared to those of the complete distributions, before the above assumption can be considered legitimate.

The assumption of a known standard deviation will, as shown below, increase the weight of the mean of the truncated distribution considerably, so that this principle should be applied as far as possible. The principle is illustrated by the following example.

Example 1. Various doses of tetanus toxin 2C were given intra-

venously to groups of 10 mice each (Ipsen, 1941, protocol 103).[1] The death times for the mice (ranging from 17 to 146 hours) were transformed into normally distributed reaction metameters (y) which follow a function that contains the reciprocal of the death time, being 0 at "infinite" death time. Table 3 presents these transformed death times for each dose of toxin.

TABLE 3

Survival of mice following injection of tetanus toxin

$x = \log(\text{mg. of toxin}).10^4$							
2.0	1.5	1.1	0.8	0.5	0.3	0.2	0.1
>1.93	1.93	1.44	0.86	0.79	0.40	0.31	0.19
>1.93	1.93	1.21	0.86	0.61	0.39	0.31	0.19
>1.93	1.63	1.21	0.86	0.61	0.29	0.22	0.12
>1.93	1.63	1.11	0.86	0.61	0.29	0.22	
>1.93	1.44	1.11	0.86	0.49	0.22	0.22	
>1.93	1.44	0.86	0.86	0.49	0.22	0.22	
>1.93	1.44	0.86	0.86	0.46	0.22	0.15	
1.63	1.44	0.86	0.86	0.40	0.22		
1.63	1.44	0.86	0.70	0.31	0.15		
1.63	1.21	0.86	0.61	0.29	0.12		
Survivors 0	0	0	0	0	0	3	7

The experiment was designed to show that the mean y for each dose has a rectilinear regression upon the log dose with regression coefficient 1; i. e., $\bar{y} = a + x$. Accordingly the problem is to compute the mean y for each group and test the validity of the equation. The first and the last two groups are truncated. The latter two are truncated at 0.11, which corresponds to 150 hours, after which time there were still found survivors in these groups. The first group is truncated in the "upper end" at 1.93, because 7 mice were found dead after 17 hours ($y = 1.93$). It means that they died unobserved during the first night, and since the function y varies widely in the first hours, it is safest to state that the mice died before 17 hours, instead of trying to assign some death time during the night.

For reasons explained in the work from which the data are taken, we can assume that the variance of y is independent of the mean y. Thus, it is legitimate to utilize the estimate of σ from the five com-

[1] J. Ipsen, Contribution to the theory of biological standardization. Nyt Nordisk Forlag (*Copenhagen*), 1941. 256 pages.

plete distributions for the calculation of the mean of three truncated distributions (see Table 4).

TABLE 4

Mean and variance of the five complete distributions

x	$\bar{y}$	$[y^2]$	d. f.	$s_{\bar{y}}^2$	χ^2	P
1.5	1.553	.47761	9	.05307	17.53	2— 5%
1.1	1.038	.38956	9	.04328	14.29	10—20%
0.8	0.819	.07129	9	.00792	2.62	95—98%
0.5	0.506	.21204	9	.02356	7.78	50—70%
0.3	0.252	.07576	9	.00842	2.78	95—98%
		1.22626	45	$\sigma^2 = .02725$		
				$\sigma = .165$	Weight $(\bar{y}) = 10/\sigma^2 = 367$	

TABLE 5

Mean and weight of mean of the three truncated distributions

Log dose	2.0	0.2	0.1
Truncation point (r_0)	1.93	0.11	0.11
	r	r	r
	0	0.20	0.08
	0	0.20	0.08
	0	0.11	0.01
	0	0.11	0
	0	0.11	0
	0	0.11	0
	0	0.04	0
	— 0.30	0	0
	— 0.30	0	0
	— 0.30	0	0
$S(r)$	— 0.90	0.88	0.17
$\bar{r}$	— 0.090	0.088	0.017
$\bar{r}/\sigma = \rho$	— 0.545	0.533	0.103
γ	— 0.283	0.291	0.990
$\bar{m} = \bar{r} - \gamma\sigma$	— 0.043	0.040	— 0.146
$\bar{y} = \bar{m} + r_0$	1.887	0.150	— 0.036
$W(\bar{y})$	296	294	150

There is some indication of heterogeneity as indicated by the P values of the χ^2 test. Bartlett's test for homogeneity gives a P value between 1 and 2 per cent.

The mean standard deviation is now applied for all the groups. The complete distributions will have an error of the mean equal to $.165/\sqrt{n}$. The procedure applied to the truncated groups is illustrated in Table 5: (1) y is translated to r by subtraction of the truncation point; (2) the mean r is divided by σ, which ratio is taken equal to ρ; (3) Table 1 is entered at ρ to obtain γ; and (4) the mean, $\bar{m}$, is computed as indicated and reconverted to $\bar{y}$, adding the truncation point. The weight of $\bar{y}$ is calculated as $n/\sigma^2 W(\mu)$, where the relative weight $W(\mu)$ is found from Table 1 with ρ as argument.

After the computation of all the $\bar{y}$ values and their weights, the test of goodness of fit to the presumed line $\bar{y} = a + x$ is carried out by computing

$$\chi^2 = \Sigma(a - \bar{a})^2 W(\bar{y}) = 7.851$$

where a is $(\bar{y} - x)$, and $\bar{a}$ is the weighted mean of a. Since χ^2 is 7.85 with 7 degrees of freedom, the fit is quite acceptable.

THE STANDARD DEVIATION ESTIMATED FROM THE TRUNCATED DISTRIBUTION

It frequently occurs that no assumption can be made as to independence of variance and mean, or that all the distributions in the experiment are truncated. In such cases the standard deviation must be estimated from each truncated distribution.

Example 2. The age distributions of poliomyelitis cases in cities, villages and country in Vermont, 1912-1925, are published by Aycock (1928) and the data are found in Table 6.

With the given age grouping a heavy truncation takes place for frequencies under 5 years of age. In order to present the three distributions adjusted as normal truncated distributions, probit diagrams were made with the transformed values of relative frequencies plotted against various functions of the age values.

It appeared that the logarithmic transformation gave rectilinear diagrams, and accordingly this particular function of the age was used as a metameter which was likely to be normally distributed.

For the calculation of the mean and standard deviation the group value was assessed by the following steps, indicated in the first columns

of Table 6. The log age limits of each group were noted, and the mean of these limits assigned to the group. Then the truncation point 0.70 (equal to log 5 years) was subtracted to normalize the scores into r.

The truncated mean and variance were computed and their ratio, h, found. By means of Table 2 where h is the argument, the quantities η

TABLE 6

The age distribution of poliomyelitis cases in Vermont 1912-1925

(From Aycock, example 2, *Am. J. Hyg.*, 8, 35-54, Jan. 1928)

				Truncation point 0.70					
				Cases					
				Cities		Villages		Country	
Age	Log (age limit)	Log (group mean)	r	Obs.	Exp	Obs.	Exp.	Obs.	Exp.
0-5	0.70	0.70	0	122	115.5	104	100.5	99	83.2
5-9	1.00	0.85	0.15	41	47.2	86	84.8	102	124.3
10-14	1.18	1.09	0.39	15	17.2	30	40.6	73	76.9
15-19	1.30	1.24	0.54	7	7.0	25	18.9	49	38.7
20-24	1.40	1.35	0.65	5	3.8	7	10.9	19	22.7
25-29	1.48	1.44	0.74	2	2.0	8	6.1	14	12.5
30-34	1.54	1.51	0.81	3	1.0	4	3.3	12	6.7
35-39	1.60	1.57	0.87	0	0.7	4	2.4	5	4.8
40+	(1.80)	1.70	1.00	1	1.6	5	5.3	6	9.2
				196	196.0	273	272.8	379	379.0
					$\chi^2 =$	1.93	7.80		16.34
					d. f. =	4	6		7
					$\bar{r} =$	.1221	.2208		.2981
					$s^2 =$	.043353	.067881		.072331
					$h =$	2.92	1.39		.77
					$\eta =$	.2570	.4823		.6910
					$\sigma^2 =$	.169552	.140745		.104675
				Estimated population standard deviation = $\sigma =$		.412	.375		.324
					$\gamma =$	.5214	.2534		.1132
					$\bar{m} =$	— .0927	.1258		.2614
				Estimated population mean =		.6073	.8358		.9614
				Mean age =		4.05 years	6.70 years		9.15 years
				$W(\mu) =$		487	2061		3302

and γ were interpolated, and the variance of the whole distribution found as the ratio of s^2 to η.

The expected number of cases per group was found by dividing the deviate of each log age limit from the mean by σ, and the expected propor-

tion found in a table of the area and deviate of the normal distribution. The expected number of cases were then computed by multiplication of the proportions by N, the total number of cases. The χ^2-values indicate reasonably good fits of the distributions.

As a test of significance of deviation of the means of the three series we computed

$$\chi^2 = \Sigma\,(\text{mean})^2 W(\mu) - \frac{1}{\Sigma W(\mu)}\,[\Sigma(\text{mean})\,W(\mu)]^2 = 64.05$$

which is highly significant for 2 degrees of freedom. Thus, we can assume that the mean age for acquiring poliomyelitis in the particular case was increasing with decreasing density of population.

THE VARIANCE OF THE ESTIMATES OF MEAN AND STANDARD DEVIATION

The last columns of Tables 1 and 2 give the relative weight of the estimate of the mean in cases where the variance is known or unknown respectively. Reference has been made to these quantities in Examples 1 and 2 without previous account of their derivation.

The relative weights have been obtained partly by mathematical considerations and partly by experimental sampling. The case is simplest when the standard deviation is known, or can be assumed to have negligible error.

The truncated mean $\bar{r}$ of the sample and the truncated mean of the population have both a variance equal to the truncated variance divided by n:

$$V(\bar{r}) = s^2/n \text{ or } V(\rho) = \eta/n. \tag{10}$$

Since $\bar{m} = \bar{r} - \gamma\sigma$, we may equate

$$V(\bar{m}) = \sigma^2 V(\mu) + \gamma^2 V(\sigma) \tag{11}$$

where the last term is negligible in the present case. From

$$V(\mu) = V(\rho)\,[d\mu/d\rho]^2 \tag{12}$$

and

$$d\rho/d\mu = \frac{d}{d\mu}(Q\mu + z) = Q - z\mu + z\mu = Q \tag{13}$$

we obtain by insertion in (10)

$$V(\mu) = (1/n)\,(\eta/Q^2) \tag{14}$$

whence (11) becomes

$$V(\bar{m}) = (\sigma^2/n)(\eta/Q^2) + \gamma^2 V(\sigma). \tag{15}$$

The factor Q^2/η is the relative weight to be given to the mean of a truncated distribution in comparison to the mean of a complete distribution with the same n and variance. This factor is given in the last column of Table 1 as $W(\mu)$.

Equation (15) is now accepted as a working hypothesis in the case where the variance has to be estimated from the distribution itself, and where $V(\sigma)$ is an important quantity.

We have equated s^2 to $\eta' = \eta\sigma^2$, and accordingly it is assumed that

$$V(\sigma) = V(s)/\eta. \tag{16}$$

In a complete distribution the estimate of the standard deviation has a variance approximately equal to $s^2/2n$. In the truncated distribution which is becoming increasingly skew the larger the truncation is, it is logical that $V(s)$ increases by a factor which is positively correlated with the degree of truncation.

This factor was determined experimentally. A series of random distributions were truncated at various points, and the variance of the truncated mean and standard deviation and the mean and standard deviation were determined in a group of samples with the same degree of truncation.

Experiments with random truncated distributions. Random numbers were transformed into probits, so that the theoretical population had mean 5.0 and unit variance. The probit values were grouped within ranges of 0.2 unit, partly to facilitate the computations and partly because most biological data are given in grouped figures, and the experimental testing of the variance should be made similar to the conditions met in practical work.

Twenty samples of the same size were randomized for each truncation point; the sample size varied from one truncation point to another, as it will appear from Tables 7 and 8. For each distribution $\bar{r}$, s, $\bar{m}$ and σ were computed and the mean and standard deviation for each of these parameters were found for each of the groups.

Truncated mean and standard deviation. The expected truncated mean is ρ and is found from Table 1, interpolating from the μ column.

TABLE 7

Truncated mean and standard deviation; comparison of experimental and expected values (20 experiments at each truncation point)

Truncation point	n	Mean $\bar{r}$		Mean s		$V(\bar{r})$		$V(s)$		h
		Obs.	Exp. (ρ)	Obs.	Exp. ($\sqrt{\eta}$)	Obs.	Exp. (η/n)	Obs.	Exp. $\frac{\eta(h+1)}{2n}$	Exp. (κ)
3.2	100	1.8214	1.8143	.9648	.9685	.01003	.00938	.003810	.00603	.285
4.0	10	1.0590	1.0836	.8308	.8608	.05690	.07425	.026243	.06059	.632
4.2	20	.8855	.9206	.8557	.8230	.04389	.03887	.043078	.03496	.799
4.4	25	.7372	.7694	.7454	.7719	.01333	.02384	.009434	.02384	1.007
4.6	40	.6285	.6313	.7146	.7141	.01420	.01275	.007413	.01454	1.280
4.8	50	.4980	.5722	.6425	.6855	.00662	.00940	.005947	.01144	1.435
5.0	10	.3830	.3989	.5112	.5838	.03429	.03408	.035241	.05354	2.142
5.2	20	.2960	.3072	.5149	.5152	.01113	.01327	.017939	.02529	2.812
5.4	20	.2100	.2314	.3708	.4473	.01933	.01000	.033940	.02368	3.735
5.6	25	.1884	.1689	.4084	.3805	.00537	.00579	.015123	.01759	5.075
5.8	30	.1092	.1202	.2826	.3183	.00403	.00338	.012045	.01354	7.01
6.0	30	.0633	.0834	.2080	.2617	.00145	.00228	.009673	.01237	9.85
6.0	100	.0857	.0834	.2572	.2617	.00038	.00069	.001713	.00372	9.85
6.2	40	.0598	.0563	.1929	.2115	.00104	.00112	.006656	.00845	14.11
6.4	60	.0472	.0369	.1931	.1683	.00059	.00047	.007954	.00514	20.79
6.6	100	.0245	.0236	.1256	.1319	.00018	.00017	.003750	.00280	31.24
6.8	200	.0139	.0150	.1001	.1035	.000059	.000054	.002085	.00130	47.54

At truncation point 3.2, μ thus is $5.0 - 3.2 = 1.8$. By interpolation we find $\rho = 1.8143$.

The expected truncated standard deviation is $\sqrt{\eta}$, which is found from Table 2, interpolating from the ρ value. In the same way the expected value for h is found.

As shown in Table 7, the expected values tally well with the observed $\bar{r}$ and s. The variance of $\bar{r}$ was expected to be equal to η/n, and the χ^2 test shows that the observed values calculated from each group of experiments with 19 degrees of freedom do not vary significantly from the expectation.

The variance of s was anticipated to be dependent upon $\eta/2n$, multiplied by a factor which increased with increasing degree of truncation.

The quantity $(\kappa + 1)$ fulfills these requirements, starting with 1, if the distribution is not truncated.

Other possibilities were $(1 + P)$ or $1/Q$ or powers thereof. The factor $(\kappa + 1)$ seems to fit better than any of the other quantities tested. In fact, it fits so well that the expected values of $V(s)$ calculated as

$$V(s) = (\sigma^2)(\eta)(k + 1)/2n \tag{17}$$

do not deviate significantly at random or systematically from the observed experimental values of $V(s)$.

Since the experimental range is sufficiently large to cover the situations met in practice, there is no need to search for a more complicated expression, although it is admitted that such a one could be found by complicated calculus.

Variance of the mean and standard deviation of the whole distribution. The same set of experiments yielded observed values for the variance of $\bar{m}$ and σ, the two final values sought. The variances and means are presented in Table 8.

The observed means of $\bar{m}$ and σ do not differ from the expected values (5.0 and 1.0, respectively) more than their calculated errors allow for, and there is no systematic deviation with increasing degree of truncation.

The expected variance of the standard deviation can be derived from (16) and (17), which lead to

$$V(\sigma) = \frac{\sigma^2}{n} \cdot \frac{\kappa + 1}{2}. \tag{18}$$

The expected values which are listed beside the observed values of $V(\sigma)$ in Table 8 do not deviate systematically or at random to such an extent that the hypothesis brought forward in equation (18) is contradicted.

TABLE 8

Mean and standard deviation for the whole distribution; comparison of expected and experimental values

Truncation point	n	Mean (expected, 5.00)	σ (expected, 1.00)	V (mean) Obs.	V (mean) Exp.*	$V(\sigma)$ Obs.	$V(\sigma)$ Exp.**	h Exp.
3.2	100	5.0070	.9968	.01001	.01010	.004886	.006425	.285
4.0	10	4.9410	1.0464	.10141	.10603	.110527	.081600	.632
4.2	20	4.8534	1.1447	.09675	.05517	.083572	.044975	.799
4.4	25	4.9406	1.0171	.03851	.04639	.040789	.040140	1.007
4.6	40	4.9550	1.0188	.04052	.03117	.018335	.028500	1.280
4.8	50	4.9765	1.0026	.02382	.02517	.019869	.024350	1.435
5.0	10	4.9711	.9410	.18184	.16134	.140054	.157100	2.142
5.2	20	4.7932	1.1299	.25320	.09919	.120384	.09530	2.812
5.4	20	5.1176	.8111	.07231	.13035	.137649	.11838	3.735
5.6	25	4.9581	1.0802	.05897	.14840	.066005	.12150	5.075
5.8	30	5.0760	.8882	.15547	.18831	.105490	.13350	7.01
6.0	30	4.8613	.9557	.78354	.30258	.301716	.18083	9.85
6.0	100	4.9512	.9602	.11656	.09078	.046990	.05425	9.85
6.2	40	5.2792	.8100	.28501	.38093	.131874	.18888	14.11
6.4	60	4.9882	1.0486	.38460	.44433	.198839	.18158	20.79
6.6	100	5.1866	.9030	.52095	.47748	.190689	.16120	31.24
6.8	200	4.7885	1.0709	.99014	.42675	.236793	.12135	47.54

* $V(\text{mean}) = (\sigma^2/n)\,(\eta/(1-p)^2 + \gamma^2(h+1)/2)$.

** $V(\sigma) = (\sigma^2/n)\,[(h+1)/2]$.

For the expected values of $V(\mu)$ equation (15) is utilized with a substitution of the value of $V(\sigma)$ given by equation (18), so that we assume that

$$V(\mu) = \frac{\sigma^2}{n} \cdot \frac{n}{Q^2} + \gamma^2 \frac{\sigma^2(\varkappa + 1)}{2n} = \frac{\sigma^2}{n}\left[\frac{n}{Q^2} + \gamma^2 \cdot \frac{\varkappa + 1}{2}\right]. \quad (19)$$

The variances calculated according to equation (19) were again compared to the variances estimated from the experiment, and also in this case no significant random or systematical deviation was found.

Thus, it is established empirically that expressions for the variance of the standard deviation and mean, which were suggested on the basis of a few elementary assumptions, did cover the observed facts. The experimental range of the degree of truncation and the grouping of data in the experiments are of the order of magnitude which is reasonable to expect in most biological data for which this method is intended.

The proposed method may not present the most exact solution of the problem, but it is not very complicated to use, and the above presentation demonstrates that it has no systematic error and the random error has been estimated satisfactorily.

In order to facilitate the calculation of the relative weights of the estimates of the mean and standard deviation, the values of $W(\mu)$ and $W(\sigma)$ are given in Table 2 for representative values of h.

STATISTICAL METHODS FOR EVALUATION OF DIAGNOSTIC AND OTHER PROCEDURES*

I. AN OBJECTIVE WEEDING-OUT PROCESS APPLICABLE TO MATERIAL USED IN SURVEYS OF DIAGNOSTIC VARIABILITY

BY WILLIAM R. THOMPSON

Division of Laboratories and Research
New York State Department of Health, Albany

INTRODUCTION

THE question of reliability of prescribed technical procedures, such as diagnostic examinations by the same or different observers, presents insistent problems of evaluation. One plan for a survey of variability in diagnosis among independent observers of the same material is to obtain a generous unbiased sample from an appropriate source of such material, and have each observer render an opinion on each specimen. This was the plan in a recent study of diagnosis of tuberculosis from chest X-ray photographs by Birkelo, Chamberlain, Phelps, Schools, Zacks, and Yerushalmy (1). The same experience was used by Yerushalmy (2) and by Neyman (3) to illustrate certain forms of statistical analysis applicable to such data.

However, we may be confronted by a dilemma in that we do not wish to spend great effort of all observers on certain kinds of material, such as that which would nearly always yield a negative diagnosis, yet we wish to have all observers examine other materials thoroughly. If we have some one select the latter material arbitrarily, no matter how skilled he may be, there is always a question as to how much such subjective influences may affect the resultant estimates of variability in diagnosis. Indeed, it is doubtful that the results will have any precise meaning for anyone but the selector.

As an alternative to this, it might be useful to have an objective

* Presented at the meeting of the Society for the Study of Mass Medicine, New York, N. Y., May 18, 1948.

method designed to eliminate a considerable portion of the unwanted material in a manner that could be definitely recorded for any future reference. The present purpose is to describe such a method based on a system of direct probability sequential analysis that has been presented elsewhere (4), and to indicate how the consequently altered sampling conditions may be taken fully into account in analysis of the data. The intention is to suggest rather than to prescribe the manner of application.

1. OBJECTIVE WEEDING-OUT PROCESS IN GOVERNING OBSERVATIONS

Accordingly, suppose that in such a situation we have n reports on a given specimen, a of which *affirm* and b *deny* the specified diagnosis in question, and that all observers requested to do so either affirm or deny unequivocally on the basis of their independent judgments after thorough examination of the specimen ($n = a + b$). Suppose that these observers may be considered as a random sample drawn from an infinite population of prospective independent observers of the same specimen (or essentially the same evidence in each instance), and that as the number of observations n (one by each observer) is indefinitely increased, the ratio a/n may be expected to approximate an unknown value ϕ, the *true* probability of affirmative diagnosis for the hypothetical infinite population of observers with respect to the given specimen (or equivalent evidence). Then, for any number p not less than zero nor greater than unity, we should like to estimate $\bar{P}(\phi < p)$, the probability that ϕ is less than p in view of the evidence of the given n observations (a affirming and b denying). Although it is not ordinarily possible to evaluate $\bar{P}(\phi < p)$ precisely, it is possible by interpretation in terms of a problem previously solved (5) to obtain a best lower limit L, and a best upper limit H, to $\bar{P}(\phi < p)$ which are given by

$$(1) \qquad L = I_p(a+1, b) \text{ and } H = I_p(a, b+1)$$

where the I-function is the Incomplete Beta-Function Ratio tabulated by Karl Pearson (6). Correspondingly, it can be shown that $\bar{P}(\phi > p) \equiv 1 - \bar{P}(\phi < p)$ is bounded by $L' = 1 - H$ and $H' = 1 - L$ where

$$(2) \qquad L' = I_q(b+1, a), \; H' = I_q(b, a+1), \text{ and } q \equiv 1 - p.$$

Thus we have $L \leq \bar{P}(\phi < p) \leq H$ and $L' \leq \bar{P}(\phi > p) \leq H'$, as judged by the evidence of the n observations comprising a affirmations and b denials of the diagnosis in question; and $\bar{P}(\phi < p'') - \bar{P}(\phi < p') = \bar{P}(p' < \phi < p'')$, the probability that ϕ lies between p' and p''.

In an obvious manner a weeding-out process may be based on such probability estimates with a specified maximum risk in each instance of inadvertent elimination of wanted material. Accordingly, after each observation on a given specimen we note whether or not it may be excluded from further observations with a risk not exceeding a prescribed tolerance T that we would be wrong in asserting that ϕ lies in some prescribed range or ranges.

Example of a weeding-out system. In order to illustrate how such a scheme may operate, consider as an example that in a survey (possibly preliminary) we wish to avoid expenditure of great effort in examinations of specimens where ϕ should be expected to be *low* (i. e., where most diagnoses would be "negative"); and in precise terms we are ready to accept a risk not to exceed 0.05 that any given specimen so eliminated would have $\phi > 0.5$. This means that we should set aside any specimen as soon as we find that $\bar{P}(\phi > 0.5) \leq 0.05$, i. e., if $0.05 \geq H'$ which equals $I_{0.5}(b, a+1)$ in this case. Table 1 gives n_a defined as the least value of n that permits such exclusion for a given value of a, i. e., the least n such that $I_{0.5}(n-a, a+1) \leq 0.05$. The table lists corresponding values of L' and H', the best lower and upper bounds of $\bar{P}(\phi > p)$ for $p = 0.75$, 0.5, and 0.25. Of course, as the plan specifies, the values of H' given in column six (where $p = 0.5$) never exceed 0.05. Furthermore, the values of H' given in the fourth column show that the risk is always less than 0.001 that we eliminate a specimen for which the *true* probability ϕ of affirmative diagnosis exceeds 0.75. The mounting values of L' for $\bar{P}(\phi > 0.25)$ are interesting in that $1 - L' \geqq \bar{P}(\phi < 0.25)$. Obviously, the probability is even less than that that a random specimen drawn from those remaining would have $\phi < 0.25$.

Of course, the tolerance T may be taken different from 0.05, perhaps higher in a preliminary survey than in a subsequent more extensive survey. Likewise, the critical value of p (taken as 0.5 above) may be chosen differently in accord with our purpose in a given situation. The principle here involved resembles that previously discussed (7, 8) in presenting the notions of sequential analysis and associated apportionment. Not only do we make successive examinations of the data as they are obtained, but we use the evidence as it accumulates in order to avoid wasteful expenditure of experimental material (the case treated formerly (7, 8) where the *material* may be human patients) or to avoid waste of the experimenter's efforts as in the present case.

In our example weeding-out system suppose that we have 20 observers and start with 400 prepared vaginal smear slides, distributed at random, 20 slides to each observer, who is given the name and address of his successor in a circuit of observation. Each of the twenty observers examines the slides sent him, passes them on, and reports the results to the office managing the experiment. There the data are scanned, and notices are sent to discontinue observations on any slide for which a

TABLE 1

Values of L' and H', respectively the best lower and upper bounds of the probability, $\bar{P}(\phi > p)$, in view of the evidence of a affirmations and b denials of the diagnosis in question by $n = a + b$ observers independently. The listed values of n_a are the least n for the given value of a yielding a risk definitely not greater than 0.05 that ϕ exceed one-half, where ϕ is the unknown *true* probability of affirmative diagnosis.

		$\bar{P}(\phi > 0.75)$		$\bar{P}(\phi > 0.5)$		$\bar{P}(\phi > 0.25)$	
a	n_a	L'	H'	L'	H'	L'	H'
0	5	0.000 00	0.000 98	0.000 00	0.031 25	0.000 00	0.237 31
1	8	.000 01	.000 39	.003 90	.035 16	.100 11	.367 09
2	11	.000 00	.000 13	.005 85	.032 72	.197 09	.455 21
3	13	.000 01	.000 13	.011 23	.046 15	.332 60	.584 26
4	16	.000 00	.000 04	.010 63	.038 41	.404 98	.630 19
5	18	.000 00	.000 04	.015 44	.048 13	.518 66	.717 46
6	21	.000 00	.000 02	.013 30	.039 18	.566 58	.743 63
7	23	.000 00	.000 01	.017 34	.046 57	.653 72	.803 70
8	26	.000 00	.000 01	.014 47	.037 76	.685 15	.819 55
9	28	.000 00	.000 01	.017 84	.043 58	.750 14	.861 55
10	30	.000 00	.000 01	.021 38	.049 37	.803 40	.894 28

affirmations have been made in the first n_a observations. In the following discussions we shall assume that further observations either are avoided by prompt action or excluded from the subsequent main analysis of data. If only half the specimens are eliminated after five observations and a total of 3/4 after 8 observations, this would save more than half (21/40) of the total work in an alternative "round-robin" experiment, if equal effort is expended on an examination of any such slide regardless of resultant diagnosis. However, even more effort might be used in examining a "negative" specimen than otherwise; hence even greater savings might result. Moreover, much greater proportions might be eliminated in some cases.

Some greater difficulty may be involved in statistical analysis of the results than in the case of round-robin surveys; likewise there would be a greater compromise of the issue resultant upon any losses of specimens after n_0 and before all required observations had been made. Such difficulties should be weighed against possible advantages of a weeding-out process. Furthermore, we should consider the possibilities of bias of observers by knowledge that a weeding-out process is in use. To combat such tendencies, we may deliberately (with or without warning according to our purpose) leave among the specimens to be further examined some specimens taken at random from the *excluded* category. Obviously, such retention may be made in various ways, and in any case tendencies to bias may be considered part of the study.

If, indeed, the judgment of observers is subject to considerable bias by any indication with regard to the relative frequency of "positive specimens," if an affirmative diagnosis is more likely in view of the *same* evidence when the observer believes on other grounds that affirmation should be more likely; then this alone suggests that the technic, at least as employed by the observers so influenced, does not meet minimum requirements for a trustworthy diagnostic method. It is obvious that evidence presented by the ordinary round-robin survey would not be enlightening in this regard; hence a weeding-out process might be recommended, even if for nothing else, to serve as a means of checking the usual assumption of unbiased, independent observation. This assumption is necessary to legitimate application of the usual methods of statistical analysis (2, 3) and to those given in subsequent sections of the present article. Usually, a preliminary coaching and training of prospective observers is used to promote not only unbiased, independent diagnosis but also their employment as nearly as practicable of the same criteria (1, 2).

However, that fatigue may influence observation is well known. Moreover, risks of this may be greater, the greater the *dilution* of "positive" with "negative" specimens. An objective weeding-out process should aid in decreasing such influences, if thereby we obtain a wanted result with less effort spent on an undesirably great relative frequency of certain types of specimens. Good conditions of observation, including relief from fatigue, should be sought in any case, but their relative cost in round-robin surveys may be prohibitive.

Of course, we should not overlook the possibilities of comparisons with other evidence from the same source as recently exploited by Isbell (9) in a study of correlation between vaginal smear and tissue

diagnosis in 1045 operated gynecologic cases. Such comparative material may be prepared even though it is not used in the first survey. Moreover, specimens discarded systematically in accord with the suggested procedure may subsequently be examined further as the interest may appear.

In the following sections of this article the problem of estimation of the character of the sampled population is considered. The relations in Section 2 are essential; Sections 3 and 4 present only one type of approach to the evaluation problem, but this is in close agreement with methods proposed by Neyman (3, 10) as applicable to round-robin surveys. A relatively simple method applicable in either case is given in Section 5. It has the further advantage of avoiding any hypothesis about the population form. This is in accord with the fundamental notions previously developed (4, 5) that have been applied in the weeding-out process itself. The arrangement of sections has been made to facilitate reading part of the article separately by those not immediately concerned with the estimation problem.

2. PROBABILITY LINK BETWEEN WEEDING-OUT AND ROUND-ROBIN SYSTEMS

Now, we shall try to avoid being unnecessarily or prematurely specific in any hypothesis about the sampled universe U, alternatively called the *population* of specimens. We assume that the character of U though unknown is fixed by the conditions of the scientific problem; perhaps the character may be described in terms of a certain set of *undetermined parameters.* In any case, let $P'_{a,n}$ denote the probability that a random specimen from the universe U will yield in n independent observations just a affirmations. Then $P'_{a,n}$ always has a definite meaning, and may be expressible in terms of undetermined parameters of U. However, it may be necessary or convenient to know the relation between $P'_{a,n}$ and P_a, the latter being defined as the probability that, under the operation of a given *weeding-out process*, just a affirmations will result. As above, it is assumed that any given specimen is drawn at random from U, and that the weeding-out process is defined by given *fixed* values of n_a for $a = 1, \cdots, K$, where K is the greatest possible value of n according to the weeding-out scheme. Let the *cessation numbers* (n_a) be an increasing sequence for $a \leq h$ and let $n_a = K$ for $a = h + 1, \cdots, K$.

Note: In the *example weeding-out system* used in illustration, we have $n_0 = 5$, $n_1 = 8$, $n_2 = 11$, $n_3 = 13$, $n_4 = 16$, $n_5 = 18$, and $n_a = 20$ for $a > 5$. Here $h = 5$, and $K = 20$.

Now, in general form without specification of the population U, we can show how to evaluate the ratio $R_a = P_a : P'_{a,n_a}$ for *any* specified weeding-out process. For convenience, we define

$$C_a = \binom{n_a}{a} \cdot R_a = \binom{n_a}{a} \cdot P_a : P'_{a,n_a}. \tag{3}$$

Since there is no weeding out until after we have $n = n_0$, obviously $P_a = P'_{a,n_a}$ for $a = 0$, hence $R_0 \equiv 1$. Of course $\binom{n_0}{0} \equiv 1$, and therefore $C_0 \equiv 1$. It is readily verified furthermore that C_a can be obtained from the recursion formulas:

$$C_a = \binom{n_a}{a} - \sum_{t=0}^{a-1} C_t \cdot \binom{n_a - n_t}{a - t} \text{ for } 0 < a \leq h; \text{ and} \tag{4}$$

$$C_a = \binom{K}{a} - \sum_{t=0}^{h} C_t \cdot \binom{K - n_t}{a - t} \text{ for } a > h. \tag{5}$$

In Table 2 are given values of C_a and R_a thus obtained for the *example system*.

3. A MODIFIED FORM OF CHI-SQUARE FUNCTION (FOR FITTING PARAMETERS TO DATA)

Without specification as to just what, if any, weeding-out system is to be used, consider certain general relations. Let N denote the total number of specimens in our random sample from U (the population of specimens). Let N_a be the number of these for which we obtain just a affirmations in the prescribed number of observations (denoted by n_a). In a round-robin system n_a is constant; in a weeding-out process it may be made dependent upon a in a prescribed manner. As previously, let K denote the greatest value of n_a. Then, of course, we have the identity

$$\sum_{a=0}^{K} N_a \equiv N. \tag{6}$$

Now, suppose that we have some *hypothetical population* $\hat{U}$ described in terms of a finite set of undetermined parameters (θ_i) to be evaluated in some prescribed manner so that $\hat{U}$ may be taken as an estimate of the specimen population U, and that $\hat{P}'_{a,n}$ is the corresponding estimate of $P'_{a,n}$ in terms of these undetermined parameters. Likewise let $\hat{P}_a \equiv R_a \cdot \hat{P}'_{a,n_a}$ and $\hat{N}_a \equiv N_a \cdot \hat{P}_a$. We suppose that $\hat{U}$ is a possible description of U and hence that

$$\sum_{a=0}^{K} \hat{P}_a \equiv 1, \text{ hence } \sum_{a=0}^{K} \hat{N}_a \equiv N. \tag{7}$$

Now, let (W_a) be any set of non-negative real numbers such that

$$\sum_{a=0}^{K} W_a = N, \tag{8}$$

but temporarily not otherwise defined; and consider the expression,

$$G^2 = \sum_{a=0}^{K} \frac{(N_a - \hat{N}_a)^2}{W_a}. \tag{9}$$

In the special case where we take $W_a \equiv N_a$, we have the χ^2 used by

TABLE 2

Evaluation of the ratio $R_a = P_a : P'_{a,n_a}$ independently of any hypothesis about the sampled universe; dependent solely upon the set of cessation numbers (n_a), a non-decreasing sequence.

a	n_a	$\binom{n_a}{a}$	C_a	$R_a = C_a : \binom{n_a}{a}$
0	5	1	1	1.000 000
1	8	8	5	0.625 000
2	11	55	25	0.454 545
3	13	286	130	0.454 545
4	16	1 820	570	0.313 187
5	18	8 568	2 916	0.340 336
6	20	38 760	12 843	0.331 347
7	20	77 520	53 569	0.691 035
8	20	125 970	110 175	0.874 613
9	20	167 960	158 670	0.944 689
10	20	184 756	180 298	0.975 871
11	20	167 960	166 240	0.989 759
12	20	125 970	125 455	0.995 912
13	20	77 520	77 410	0.998 581
14	20	38 760	38 745	0.999 613
15	20	15 504	15 503	0.999 936
16	20	4 845	4 845	1.000 000
17	20	1 140	1 140	1.000 000
18	20	190	190	1.000 000
19	20	20	20	1.000 000
20	20	1	1	1.000 000

Neyman (3) to fit parameters of a given hypothesis to the data by minimizing G^2. However, in that case we must stipulate that N_a is nowhere zero ($N_a \neq 0$ for every a), else no solution is possible. We shall exclude the possibility that $P_a = 0$ for any a; hence, as $\hat{U}$ is a possible form of U, we have

$$\hat{N}_a \equiv N \cdot \hat{P}_a > 0 \text{ for every } a. \tag{10}$$

In order to avoid the difficulty, we may take W_a as a non-decreasing linear function of N_a, with W_a having a fixed value $w > 0$ when $N_a = 0$. Then, with extension for the sake of generality in formulation to include the case $w = 0$, it is readily shown to be necessary and sufficient that

$$W_a = \frac{N - (K+1)w}{N} \cdot N_a + w \tag{11}$$

provided that $N - (K+1)w \geq 0$.

Correspondingly let G_w denote G of relation (9) with W_a defined by (11). Then $G_0 = \chi$. Now, let g be defined by

$$g = \frac{N \cdot w}{N - w(K+1)}. \tag{12}$$

Obviously, for $w > 0$, we have $g > 0$ and

$$G^2{}_w = \frac{g}{w} \cdot \sum_{a=0}^{K} \frac{(N_a - \hat{N}_a)^2}{N_a + g}. \tag{13}$$

Let G denote G_1 unless otherwise specified. Then we may write simply

$$G^2 = g \cdot \sum_{a=0}^{K} \frac{(N_a - \hat{N}_a)^2}{N_a + g} \tag{14}$$

where $g = \dfrac{N}{N - K - 1} > 1$.

Now, it may be convenient to let the index i, of our set (θ_i) of undetermined parameters of $\hat{U}$, be defined as running from $-u$ to v inclusive, i.e., $i = -u, \cdots, v$. Then the number of undetermined parameters is $u + v + 1$. This may be convenient as seen in Section 4 below. If we are to evaluate these parameters by means of minimizing G^2, then we must set the partial derivatives $\partial G^2/\partial\theta_i = 0$ for every i. Thus, in (14) we obtain the relations:

$$\partial G^2/\partial\theta_i = 2g \cdot \sum_{a=0}^{K} \frac{N_a - \hat{N}_a}{N_a + g} \cdot \frac{\partial \hat{N}_a}{\partial \theta_i}, \tag{15}$$

where by (7) we have

$$\sum_{a=0}^{K} (\partial \hat{N}_a / \partial \theta_i) \equiv 0. \tag{16}$$

Accordingly, relation (15) may be simplified by use of (6), (7), and (16) to eliminate N_a, $\hat{N}_a$, and $\partial \hat{N}_a / \partial \theta_i$ for any one value of a, say $a = 0$.

4. EVALUATION OF PARAMETERS OF A HYPOTHETICAL UNIVERSE (SAMPLED POPULATION) OF GIVEN CATEGORICAL FORM

Suppose that for a certain finite set of values (p_j), where $0 \leq p_j \leq 1$ and $j = 0, \cdots, u$, that $P(\phi = p_j)$ is *not necessarily zero*, but that for any $p \neq p_j$ for every j we have $P(\phi = p) \equiv 0$. Then, let

$$F_{-j} \equiv P(\phi = p_j) \tag{17}$$

be the first $u + 1$ of the undetermined parameters (θ_i), respectively, for $i = -u, \cdots, 0$. Let us suppose, however, that unity and zero are the first and last values of p_j, i. e., $p_0 = 0$ and $p_u = 1$. These are the most desirable values of ϕ, of course, since *all* observers (no matter how many) are expected to agree on the diagnosis of specimens whose $\phi = 0$ or 1. Since it seems unlikely that we should want to postulate that $P(\phi = p) \neq 0$ unless $p = 0$ or 1, let us suppose, to simplify the present discussion, that there are only these two values of (p_j), i. e., $p_0 = 0$ and $p_1 = 1$, and accordingly $u = 1$. Then, the parameters (F_{-j}) consist of $F_{-1} = P(\phi = 1)$ and $F_0 = P(\phi = 0)$.

Now, let the unit interval $(0, 1)$ be divided into a set of v successive touching intervals (J_i) of length Δ_i, i. e.,

$$J_i = (p'_i, p''_i) \text{ and } \Delta_i = p''_i - p'_i \tag{18}$$

for $i = 1, \cdots, v$, where $p'_{i+1} = p''_i$, $p'_1 = 0$ and $p''_v = 1$.

Then, let F_i be the relative frequency in $\hat{U}$ of ϕ *within* J_i, exclusive of the values (p_j), namely 0 and 1. For simplicity we may consider p'_i and p''_i excluded also, since this would not affect the value of F_i because we have stipulated that $P(\phi = p) \equiv 0$ if p is not in the set (p_j). Similarly let $f(p) \cdot dp$ denote the relative frequency of ϕ within the interval $(p, p + dp)$, which, of course, is in reference to the *true* universe U. Correspondingly, let $\hat{f}(p)$ instead of $f(p)$ apply to the hypothetical universe $\hat{U}$ with which we aim to approximate U. Again for simplicity, let us stipulate that $\hat{f}(p)$ shall be constant within each of the specified intervals (J_i), and define:

$$f_i = \hat{f}(p) \text{ for } p \text{ in } J_i. \tag{19}$$

Obviously, then

$$F_i = \Delta_i \cdot f_i = (p''_i - p'_i) \cdot f_i = \int_{p'_i}^{p''_i} \hat{f}(p) \cdot dp, \tag{20}$$

for $i = 1, \cdots, v$, and

$$\sum_{i=-1}^{v} F_i \equiv 1 \tag{21}$$

where F_{-1} and F_0 are as previously defined. We have thus $v + 2$ parameters (F) which we propose to determine so as to minimize G^2, which we shall consider to be as defined in relations (14) to (16), i. e., with $w = 1$ and $G = G_1$. The result may be represented as shown in a factitious illustration in Fig. 1, where rectangles on the intervals J_i have areas proportional to F_i and a triangle with one side on the vertical at p_j and apex on the abscissal axis has an area likewise proportional to F_{-j} (where, of course, $j = 0$ or 1).

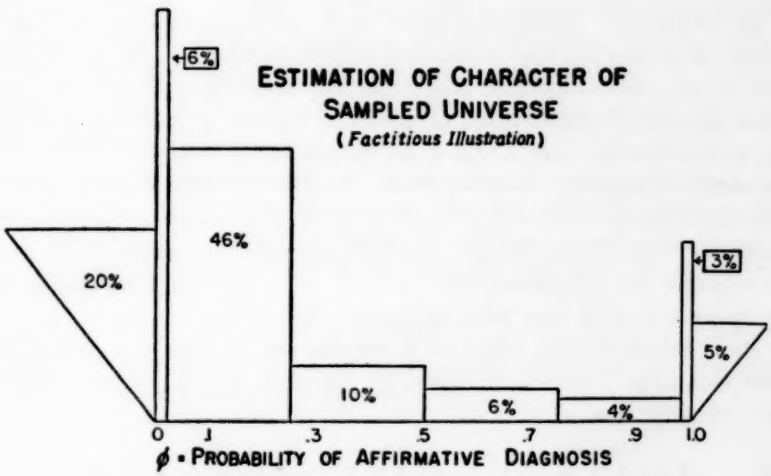

FIG. 1. ESTIMATION OF CHARACTER OF SAMPLED UNIVERSE (FACTITIOUS ILLUSTRATION)

Use of a hypothetical universe of this sort permits a number of simplifications in the technical work of estimation of the parameters, and is in approximate accord with suggestions made by Neyman (3, 10) in introducing the model system he used in analysis of the data of Birkelo,

Chamberlain, Phelps, Zacks, and Yerushalmy (1, 2). Neyman (3) pointed out that the model scheme was a simplification of the actual situation which might better be represented by a subdivision much finer than that which he used.

A number of technical economies have been developed for application of the scheme at present suggested, but will not be detailed here. It should be emphasized that χ^2 (which equals G^2_0) may be used instead of G^2_1 as above, if we are willing to take the risk that no solution will result (which is an unfortunate necessary condition if any N_a is zero in the observed results). However, it would introduce some bias if we reject such experiences or resort to use of G_w only in such an eventuality. Likewise, it should be noted that the *weeding-out scheme* of observation is a *generalization including the round-robin scheme* as a special case (where we take all $n_a = K$, a constant).

5. ALTERNATIVE (NON-PARAMETRIC) METHOD OF EVALUATION

As mentioned previously, there is a relatively simple alternative method of estimating the cumulative distribution of ϕ in the population. This is non-parametric and free from any assumption about the form of the population distribution of ϕ except that the cumulative distribution is continuous. Let $P_U(\phi < p)$ denote this cumulative probability, the relative frequency of ϕ less than p in the universe U. We proceed to estimate its expected value $\bar{P}_U(\phi < p)$ in view of the evidence of our experience with N specimens, N_a of which yielded a affirmations in n_a observations (as prescribed for $a = 0, 1, \cdots, K$), the round-robin system being a special case wherein $n_a = K$ for every a.

Now, for any given value of a, we can by relation (1) obtain best lower and upper bounds to $\bar{P}(\phi < p)$ for each of the N_a specimens in that category; i. e.,

$$(22) \qquad I_p(a+1, n_a - a) \leqq \bar{P}(\phi < p) \leqq I_p(a, n_a - a + 1)$$

since $b = n_a - a$ in this case. Furthermore, the probability that a random specimen from our sample experience of N specimens will be one of these is just N_a/N. Accordingly, we can obtain best lower and upper bounds, L_p and H_p, for the expected probability that a random specimen from the N under observation would have a value of ϕ less than a given p. Obviously, these are given by

$$L_p = (1/N) \cdot \sum_{a=0}^{K} N_a \cdot I_p(a+1, n_a - a), \text{ and} \tag{23}$$

$$H_p = (1/N) \cdot \sum_{a=0}^{K} N_a \cdot I_p(a, n_a - a + 1). \tag{24}$$

These values may likewise be taken as best lower and upper bounds to the expected cumulative distribution $\bar{P}_U(\phi < p)$ for the universe U. Much less computational difficulty is encountered than in any chi-square solution (as in Sections 3 and 4), and there is no added assumption involved with regard to a hypothetical population $\hat{U}$. The bounds may be far apart where there is considerable influence by small values of n. However, the main purpose of a weeding-out process is to decrease such influences in the regions regarded by the investigators as most important to their purposes, accomplishing this with due attention to economy. It may be convenient to define a *mid-expectation* estimate of $\bar{P}_U(\phi < p)$ as $(L_p + H_p)/2$.

An example is appended to illustrate the evaluation and relative economy with artificial data, taken roughly to represent a situation such as might have been encountered in an actual experience like that of Birkelo and associates (1) but with a greater number of observers available—contrasting the results obtained with round-robin and weeding-out processes.

Note: It is noteworthy that the Bayes-rule estimate of $P(\phi < p)$ for a specimen with a affirmations in n_a observations would be $I_p(a+1, n_a - a + 1)$. This is readily seen, from previously developed recursion formulas (11, 8) for the incomplete beta-function ratio, to be related to our best lower and upper bounds by the identity:

$$I_p(a+1, n-a+1) \equiv (1-p) \cdot I_p(a+1, n-a) + p \cdot I_p(a, n-a+1). \tag{25}$$

Accordingly, since the coefficients of the I-function values here are independent of a and n, we could obtain a *Bayes-rule* estimate of $\bar{P}_U(\phi < p)$ by

$$B_p = (1/N) \cdot \sum_{a=0}^{K} N_a \cdot I_p(a+1, n_a - a + 1), \text{ whence} \tag{26}$$

$$B_p \equiv (1-p) \cdot L_p + p \cdot H_p. \tag{27}$$

6. EXAMPLES OF NON-PARAMETRIC ESTIMATION OF CUMULATIVE FREQUENCY DISTRIBUTION OF ϕ IN ROUND-ROBIN AND WEEDING-OUT SURVEYS

As an illustration consider the partly artificial data in Table 3: I. A simulated round-robin experience of 20 observations on each of 1256 specimens. II. A part of experience I obtained by taking note

TABLE 3

Illustrative data, partly artificial, from a round-robin simulated experience I with $n_a = K = 20$ and from a corresponding weeding-out experience II with n_a as given; III is an actual experience in diagnosis of tuberculosis with large celluloid x-ray photographs[1] with $n_a = 5$ throughout.

	I		II		III	
a	N_a	n_a	N_a	n_a	N_a	n_a
0	994	20	1101	5	1125	5
1	93	"	46	8	47	"
2	45	"	13	11	23	"
3	14	"	9	13	17	"
4	10	"	2	16	17	"
5	9	"	0	18	27	"
6	7	"	3	20		
7	4	"	3	"		
8	6	"	6	"		
9	7	"	7	"		
10	5	"	4	"		
11	3	"	3	"		
12	4	"	4	"		
13	5	"	5	"		
14	6	"	6	"		
15	9	"	9	"		
16	5	"	5	"		
17	3	"	3	"		
18	3	"	3	"		
19	3	"	3	"		
20	21	"	21	"		

Note: II was obtained from the artificial data of I by the weeding-out process. However, III was not so obtained but was an actual experience with $n = 5$ throughout. The total "observations" made were 4 times as many in I as in III, but only slightly more than 25 per cent more observations were required for II than for III. As previously, a is the number of affirmations in n_a observations and N_a is the number of specimens in the corresponding category.

as to which specimens had no affirmations in the first five observations, one affirmation in five but no more in eight observations, etc., to yield the data for a corresponding weeding-out process employing the values of n_a given in Table 2. Thus from the supposed round-robin experience

I in Table 3, the weeding-out experience II was obtained. III. An actual experience of Birkelo, *et al.* (1) with a round-robin of 5 observations on each specimen.

TABLE 4

Evaluation of L_p and H_p, best lower and upper bounds to $\bar{P}_U(\phi < p)$, from the data of I, II and III, respectively, from Table 3

	I		II		III	
p	L_p	H_p	L_p	H_p	L_p	H_p
0.01	0.1454	0.8055	0.0431	0.8795	0.0439	0.8976
.02	.2678	.8182	.0846	.8823	.0862	.8994
.03	.3707	.8297	.1247	.8850	.1269	.9011
.05	.5302	.8494	.2006	.8902	.2035	.9046
.10	.7537	.8832	.3670	.9017	.3700	.9126
.20	.8912	.9157	.6140	.9192	.6132	.9266
.30	.9228	.9307	.7710	.9315	.7663	.9381
.40	.9359	.9409	.8647	.9412	.8580	.9480
.50	.9452	.9494	.9168	.9496	.9102	.9568
.60	.9533	.9574	.9442	.9575	.9388	.9649
.70	.9615	.9659	.9594	.9659	.9545	.9729
.80	.9704	.9745	.9701	.9745	.9642	.9811
.90	.9781	.9825	.9780	.9825	.9716	.9900
.95	.9809	.9884	.9809	.9884	.9751	.9948
.97	.9818	.9920	.9818	.9920	.9764	.9968
.98	.9823	.9943	.9823	.9943	.9771	.9979
.99	.9828	.9969	.9828	.9969	.9778	.9989

In Table 4 are given the corresponding estimates of L_p and H_p for I, II, and III with $p = 0.01$, 0.02, 0.03, 0.05, 0.1, 0.2, $\cdots$ 0.9, 0.95, 0.97, 0.98, and 0.99. These were computed with the aid of Karl Pearson's tables (6) and relations (23) and (24). The results are shown graphically in Figs. 2 and 3.

What forms of presentation and analysis of results should be used will largely depend on the interest of the investigators and economic factors. The main purpose here is not to indicate preferred forms of analysis but rather to point the way to economical means of obtaining the data pertinent to a given investigation. Perhaps it would be adequate in some situations merely to exhibit the results as in Table 3, where it should be noted that the added information in the weeding-out

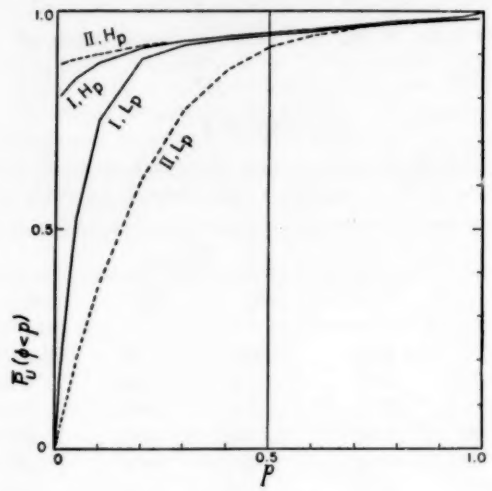

FIG. 2.

BEST UPPER AND LOWER BOUNDS, L_p AND H_p, FOR THE EXPECTED RELATIVE FREQUENCY OF ENCOUNTERING IN THE POPULATION A SPECIMEN WHOSE ϕ (PROBABILITY OF AFFIRMATIVE DIAGNOSIS) IS LESS THAN A GIVEN p.

I is according to the artificial data of a round-robin of 20 observations on each specimen; II, according to the corresponding data obtained in a weeding-out process (mean number of observations about 6.26); III is according to a round-robin of 5 observations (shown in Fig. 3).

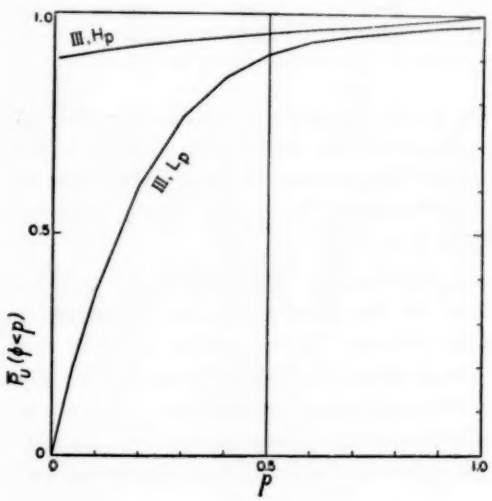

FIG. 3.

L_p AND H_p FOR III.
(Compare with I and II in Fig. 2)

survey II is obtained by an increase of about 25 per cent in the number of observations required for a round-robin survey with $n = 5$ as in III. Of course, there is no restriction to use of the form of weeding-out process used in illustration. If we can use more than 20 observers, then one recourse is to have n_a as in Table 1 for 30 observers ($K = 30$) or more. Relative economy in obtaining such additional information would be even greater than in the illustrative example. From some practical points of view it might be more appropriate to judge the variability of a diagnostic method or other technic as in actual usage by many observers rather than by a selected few. Moreover, with the suggested scheme there would always be less work expected per observer the greater their number.

Practical needs thus far unsatisfied in statistical methods for evaluation of diagnostic procedures were discussed at a conference with Dr. Gilbert Dalldorf, Dr. John K. Miller, and Dr. Victor N. Tompkins. The author wishes to acknowledge their contribution to the basic principles on which the present suggestions rest.

REFERENCES

1. Birkelo, C. C., Chamberlain, W. E., Phelps, P. E., Zacks, D., and Yerushalmy, J. Tuberculosis case finding. *J. Amer. Med. Assoc.*, 1947, 133, 359-365.
2. Yerushalmy, J. Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Pub. Health Reports*, 1947, 62, 1432-1449.
3. Neyman, J. Outline of statistical treatment of the problem of diagnosis. *Pub. Health Reports*, 1947, 62, 1449-1456.
4. Thompson, W. R. Direct probability sequential analysis. Preliminary report. Read by title at the Annual Meeting of the Amer. Math. Soc., Athens, Georgia, December 29-31, 1947; *Bull. Amer. Math. Soc.*, 1948, 54, 288-289.
5. Thompson, W. R. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. Math. Statis.*, 1936, 7, 122-128.
6. Pearson, K. Tables of the Incomplete Beta-Function. 1934, University Press (*Cambridge, England*), lix + 494 p.
7. Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 1933, 25, 285-294.

8. THOMPSON, W. R. On the theory of apportionment. *Amer. Jour. Math.*, 1935, 57, 450-456.

9. ISBELL, N. P. Correlation between vaginal smear and tissue diagnosis in 1,045 operated gynecologic cases. *Amer. J. Obs. and Gynec.*, 1947, 54, 576-583.

10. NEYMAN, J. Contribution to the theory of the χ^2-test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability of 1945-46*, Univ. of Calif. Press (*Berkeley*), 239-273. (In advance of publication, Professor Neyman kindly gave the present writer a proof of his article.)

11. THOMPSON, W. R. Recursion formulas for the tetradite ψ and the incomplete I-function. *Bull. Amer. Math. Soc.*, 1934, 40, 43.

THE EFFECT OF THE CHANGE IN MORTALITY CONDITIONS IN AN AGE GROUP ON THE EXPECTATION OF LIFE AT BIRTH

BY C. CHANDRA SEKAR *

Population Division, United Nations, Lake Success

INTRODUCTION

WHEN life tables are prepared periodically it becomes a matter of interest to know the extent to which the expectation of life at birth has been influenced by the changes in the mortality conditions in the different age groups. For instance, the rather frequent experience of a simultaneous increase in the expectation of life at birth and the marked diminution in the infant mortality rate, raises the important question as to what contribution the reduction in the infant mortality rate has made on the expectation of life at birth. Similarly the effect on the expectation of life of the improvement in mortality conditions which might have been observed in a particular age group as a result of specialised public health measures, is not without interest. In contexts such as these the expressions, "the effect of," "the influence of," or "the contribution made by" the change in mortality conditions are used synonymously. But what exactly is connoted by these expressions requires some consideration. In this article the difficulties that arise in giving a single definition to what is implied by these terms will first be pointed out, and will be followed by a discussion of the formulae and methods to evaluate the "effect," using different definitions. Attention will be mainly confined to working out the "effect" of the change in mortality conditions in an age group on the expectation of life at birth. By a logical extension of this work the "effect" of the change in mortality conditions of two or more age groups on the expectation of life at birth or at other ages could be found.

* On leave from the All-India Institute of Hygiene and Public Health.

TWO CONCEPTS OF "EFFECT"

If e_0 is the expectation of life at birth in the first period and e'_0 the expectation of life at birth in the second period, the effect of the change during this period in mortality conditions in an age group on the expectation of life at birth may be conceived in the following two ways:

(1) The difference between e_0 and the expectation of life at birth $e_{0/a}$ (say) which would have resulted if the mortality conditions had changed only in the age group under consideration to the extent that it had, and the mortality conditions in the other age groups had remained unchanged. For uniformity, the effect may be considered as $e_{0/a} - e_0$.

(2) The difference between e'_0 and the expectation of life at birth $e'_{0/a}$ (say) which would have resulted if the mortality conditions had remained unchanged in the age group under consideration, and the mortality conditions in all the other age groups had changed to the extent they had. For uniformity, the effect may be considered as $e'_0 - e'_{0/a}$.

The effect as obtained from the first consideration may be defined as the "main effect" and that obtained from the second as the "operative effect." The above concepts of the "main effect" and "operative effect" are capable of obvious extension when the influence of the change in mortality conditions of two or more age groups on the expectation of life at birth or at any other age is being considered.

RELATIONSHIP BETWEEN MAIN EFFECT AND OPERATIVE EFFECT

Notation:

In this section and all through this paper the following notation will be maintained:

e_x = expectation of life at x completed years of age in the first period,

l_x = number of survivors at x completed years of age in the first period,

e'_x = expectation of life at x completed years of age in the second period,

l'_x = number of survivors at x completed years of age in the second period

$l_0 = l'_0$ = radix of the life table.

Age span divided into two age groups

The relationship between main effect and operative effect can be readily seen by dividing the age span 0 to w years (say) into the two

groups 0 to x years and x to w years and working out the effects[1] for the two age groups.

Main effect. The change in the mortality conditions in the 0 to x age group will lead to a difference in the "number of years lived" between 0 to x years of

$$(l'_0e'_0 - l'_xe'_x) - (l_0e_0 - l_xe_x), \tag{1}$$

and between x to w years of

$$(l'_x - l_x)e_x. \tag{2}$$

The main effect of the 0 to x age group is therefore equal to

$$(1/l_0)\{(l'_0e'_0 - l'_xe'_x) - (l_0e_0 - l_xe_x) + (l'_x - l_x)e_x\}. \tag{3}$$

The main effect of the age group x to w years is

$$(l_x/l_0)(e'_x - e_x). \tag{4}$$

The sum of the two main effects differs from the difference of the expectation of life in the two periods, viz. $(e'_0 - e_0)$ by

$$(1/l_0)(l'_x - l_x)(e'_x - e_x). \tag{5}$$

Apart from l_0, this expression is made up of two other factors, one, $(l'_x - l_x)$, representing the difference in the number of survivors at age x caused by mortality changes in the age groups 0 to x years and the other, $(e'_x - e_x)$, the difference in the expectation of life at age x caused by mortality changes in the age group x to w years.

Whereas each of the main effects is solely due to the change in the mortality conditions in the age group to which it has reference, and vanishes if the mortality condition in that age group remains unchanged, the above expression is influenced by the change in the mortality conditions in either age group and vanishes if the mortality conditions in either of the age groups remain unchanged.

The influence of the "interaction" is ignored when only the "main effects" are taken into account.

[1] In calculating these effects, use will be made constantly of the concept of the number of years lived after age x and its relationship to expectation of life at age x given by,

$$\text{expectation of life at age } x = \frac{\text{number of years lived after age } x}{\text{number of survivors at age } x}$$

Operative effect. If the mortality conditions had remained unchanged in the 0 to x year group (and changed in the x to w age group to the extent it had), the "number of years lived" between 0 to x years, viz. $l'_0e'_0 - l'_xe'_x$ as given by the second life table, would have been $(l_0e_0 - l_xe_x)$. Also the "number of years lived" between x and w years would be $l_xe'_x$ instead of $l'_xe'_x$.

The operative effect of 0 to x years is therefore

$$(1/l_0)\{(l'_0e'_0 - l'_xe'_x) - (l_0e_0 - l_xe_x) \qquad (l'_x - l_x)e'_x\}. \tag{6}$$

The operative effect of x to w years is equal to

$$(l'_x/l_0)(e'_x - e_x). \tag{7}$$

The operative effect of the 0 to x age group differs from the main effect of that age group by $(1/l_0)(l'_x - l_x)(e'_x - e_x)$, which, as noticed earlier, is the "interaction" of the two age groups 0 to x and x to w years. Similarly the operative effect of the age group x to w years differs from the main effect of that age group by an amount equal to the same interaction. It is therefore seen that whereas the main effect ignores the effect of interaction, each of the "operative effects" takes it into account. For this reason, whereas the sum of the two main effects falls short of the difference in the expectation of life in the two periods by an amount equal to the interaction, the sum of the two operative effects exceeds the difference in the expectation of life by an amount equal to the interaction.

Age span divided into three or more groups

The concept of interaction introduced above in the case when the age span was divided into two age groups is capable of extension. When the age span is divided into three or more age groups and when so generalised, the interaction can be shown to be the factor producing the difference between the main effect and operative effect. The definition of interaction when more than two age groups are involved, and the nature of the mathematical relationship by which the interactions relate the main effect and operative effect, are considered below with reference to the case of three age groups.

Interaction generalised. Consider the age span 0 to w years divided into the three age groups 0 to x, x to $x + n$ and $x + n$ to w.

Let M_1, M_2, M_3 denote respectively the main effects over the three age groups,

M_{12}, the main effect over the age groups 0 to x and x to $x + n$,

M_{23}, the main effect over the age groups x to $x + n$ and $x + n$ to w,

M_{13}, the main effect over the age groups 0 to x and $x + n$ to w, and

M_{123}, the main effect over the three age groups 0 to x, x to $x + n$ and $x + n$ to w years.

The difference between the sum of the main effects over each of the two age groups and the main effect over the two age groups taken together, will be defined as the first order interaction. Denoting by I_{12}, I_{23} and I_{13} the three first order interactions possible in this instance, they are, by definition, given by

$$\left.\begin{aligned} I_{12} &= M_{12} - M_1 - M_2 \\ I_{23} &= M_{23} - M_2 - M_3 \\ I_{13} &= M_{13} - M_1 - M_3. \end{aligned}\right\} \quad (8)$$

The second order interaction will be defined as the difference between the main effect over the three age groups and the sum of the three main effects over each of the age groups plus the three first order interactions each taken over two of the three age groups. If I_{123} defines the only second order interaction possible in this case,

$$I_{123} = M_{123} - M_1 - M_2 - M_3 - I_{12} - I_{23} - I_{13}. \quad (9)$$

It will be observed that "interactions" as defined above[2] are but logical extensions of the basic concept introduced in the case of two age groups. It can also be shown that the first and second order interactions do, as their titles imply, measure influences resulting from the changes in mortality conditions of two and three age groups respectively, and vanish if the mortality conditions in any one of these age groups remain

[2] Those familiar with the set-up of factorial experiments in modern statistical designs, and the techniques of the analysis of variance for interpretation of data, will observe the analogy between the relationship of main effects and interactions introduced therein and in the present discussion. The case of three age groups is very similar to a $2 \times 2 \times 2$ factorial design. See Fisher, R. A., The Design of Experiments. Oliver and Boyd (*London*).

unchanged. For instance, in the example under discussion the various main effects are as follows:

$$\left.\begin{aligned}
M_1 &= (1/l_0)\,(l_0/l'_0)\,[l'_0(e'_0 - e_0) - l'_x\,(e'_x - e_x)] \\
M_2 &= (1/l_0)\,(l_x/l'_x)\,[l'_x(e'_x - e_x) - l'_{x+n}(e'_{x+n} - e_{x+n})] \\
M_3 &= (1/l_0)\,(l_{x+n}/l'_{x+n})\;[l'_{x+n}(e'_{x+n} - e_{x+n})] \\
M_{12} &= (1/l_0)\,(l_0/l'_0)\,[l'_0(e'_0 - e_0) - l'_{x+n}(e'_{x+n} - e_{x+n})] \\
M_{23} &= (1/l_0)\,(l_x/l'_x)\,[l'_x(e'_x - e_x)] \\
M_{13} &= (1/l_0)\,[l_0(e'_0 - e_0) - l'_x(e'_x - e_x) + \\
&\qquad (l'_x/l_x)l_{x+n}(e'_{x+n} - e_{x+n})] \\
M_{123} &= (1/l_0)\,[l'_0e'_0 - l_0e_0]
\end{aligned}\right\} \qquad (10)$$

Applying (8) and (9) to obtain the values of the interactions,

$$\left.\begin{aligned}
I_{12} &= (1/l_0)[(l_0/l'_0) - (l_x/l'_x)]\,[l'_x(e'_x - e_x) - l'_{x+n}(e'_{x+n} - e_{x+n})] \\
I_{23} &= (1/l_0)[(l_x/l'_x) - (l_{x+n}/l'_{x+n})]\;l'_{x+n}(e'_{x+n} - e_{x+n}) \\
I_{13} &= (1/l_0)\,l_{x+n}[(l'_x/l_x) - 1]\,(e'_{x+n} - e_{x+n}) \\
I_{123} &= (1/l_0)(l_x - l'_x)[(l_{x+n}/l_x) - (l'_{x+n}/l'_x)]\,(e'_{x+n} - e_{x+n})
\end{aligned}\right\} \qquad (11)$$

The formulae for the interactions demonstrate clearly that they do measure contributions to the change in the expectation of life at birth resulting from the changes in mortality conditions in two or three age groups as the case may be.

Operative effect in terms of main effect and interactions

Starting from definition, the "Operative effect" over each of the three age groups can be worked out. If O_1, O_2 and O_3 denote respectively the operative effects over the age groups 0 to x, x to $x+n$ and $x+n$ to w years, it can be shown that

$$\left.\begin{aligned}
O_1 &= (1/l'_0)\,(l'_0/l_0)\,[l_0(e'_0 - e_0) - l_x(e'_x - e_x)] \\
O_2 &= (1/l'_0)\,(l'_x/l_x)\,[l_x(e'_x - e_x) - l_{x+n}(e'_{x+n} - e_{x+n})] \\
O_3 &= (1/l'_0)\,(l'_{x+n}/l_{x+n})\,[l_{x+n}(e'_{x+n} - e_{x+n})]
\end{aligned}\right\} \qquad (12)$$

Comparing the formulae (10), (11) and (12), it is seen that

$$\left.\begin{aligned} O_1 &= M_1 + I_{12} + I_{13} + I_{123} \\ O_2 &= M_2 + I_{12} + I_{23} + I_{123} \\ O_3 &= M_3 + I_{13} + I_{23} + I_{123}. \end{aligned}\right\} \quad (13)$$

Again it is noticed that the operative effect of an age group is equal to the sum of the main effect of the age group and all the interactions in which that age group plays a part.

DISCUSSION

When the age span is divided into a number of age groups, the difference in the expectation of life at birth at any two periods is equal to the sum of the main effects over the different age groups plus all the interactions between these age groups. The interactions, resulting as they do from the change in mortality conditions in two or more age groups, are not wholly attributable to any single age group. In considering therefore the main effect as the contribution of the change in mortality conditions of an age group on the expectation of life at birth, (11) the part that this age group contributes toward interaction is ignored. On the contrary if the "operative effect" is considered as the contribution of the age group, that age group is given the onus of all the interactions toward whose being that age group is only partly responsible. The absence of a logical basis for apportioning the interaction between the different age groups is the main difficulty in working out the contribution of the different age groups to the change in the expectation of life at birth.

One method of getting over this difficulty is to evaluate both the main effect and operative effect and consider these as the two limits within which the value of the contribution lies. Another is to formulate some arbitrary procedure by which the interaction could be assigned to one or other of the age groups responsible for producing it. For instance it may be decided to assign an interaction to the oldest age group involved in its production. Alternatively it may also be decided to assign an interaction to the youngest age group concerned. If the contributions of the different age groups worked out on these two assumptions may for brevity be called "Effect-interaction deferred"[3] and "Effect-

[3] This procedure was used by the author in "Some aspects of Parsi demography," *Human Biology*, Vol. 20, p. 77, 1949.

interaction forwarded" respectively, it will be observed that the sum of the "effects" over all the age groups obtained by either method will exactly account for the difference in the expectation of life at birth in the two periods. These methods of evaluating the effect have also the advantage, which the concepts of "main effect" and "operative effect" do not possess, that the effect over two or more adjoining age groups is easily obtained by adding the effects calculated over each of the age groups. The nature of the formulae for obtaining the two effects—with interactions deferred and forwarded—can be seen by working these out for the case considered previously, *viz.*, of the age span divided into three groups. These are as follows:

Effect-interaction deferred

$$\left.\begin{aligned}
&\text{Effect of age group 0 to } x = \\
&\qquad M_1 = (1/l'_0)\,[l'_0(e'_0 - e_0) - l'_x(e'_x - e_x)] \\
&\text{Effect of age group } x \text{ to } x+n = \\
&\qquad M_2 + I_{12} = (1/l'_0)\,[l'_x(e'_x - e_x) - l'_{x+n}(e'_{x+n} - e_{x+n})] \\
&\text{Effect of age group } x+n \text{ to } w = \\
&\qquad M_3 + I_{13} + I_{23} + I_{123} = (1/l'_0)\,[l'_{x+n}(e'_{x+n} - e_x)]
\end{aligned}\right\} \quad (14)$$

Effect-interaction forwarded

$$\left.\begin{aligned}
&\text{Effect of age group 0 to } x = \\
&\qquad M_1 + I_{12} + I_{13} + I_{123} = \\
&\qquad\qquad (1/l_0)[l_0(e'_0 - e_0) - l_x(e'_x - e_x)] \\
&\text{Effect of age group } x \text{ to } x+n = \\
&\qquad M_2 + I_{23} = (1/l_0)\,[l_x(e'_x - e_x) - l_{x+n}(e'_{x+n} - e_{x+n})] \\
&\text{Effect of age group } x+n \text{ to } w = \\
&\qquad M_3 = (1/l_0)\,[l_{x+n}(e'_{x+n} - e_{x+n})],
\end{aligned}\right\} \quad (15)$$

In fact the effect for age group x to $x+n$ given above can be used more generally to calculate the effect of any age group, irrespective of the number of groups into which the age span is divided.

It does not appear possible to advocate any one procedure of allocating the interaction as superior from a theoretical point of view, and the method of choice in any particular case may therefore have to rest

on the practical issues involved. An appreciation of the precise nature of the difficulties, as set forth above, that arise in evaluating the effect should undoubtedly assist in the choice of the method to be used in any particular instance.

Illustrations.[4] The formulae discussed in the foregoing text are applied below to the data provided by the life tables for males in Sweden in the two periods 1901-1910 and 1936-1940. Table 1 shows the number of survivors and the expectation of life at different ages as given by the two life tables.

TABLE 1

Number of survivors and expectation of life at different ages in Swedish life tables for males 1901-10 and 1936-40

Age	*1901-1910**		*1936-1940***	
	Number of survivors	Expectation of life	Number of survivors	Expectation of life
0	100 000	54.53	100 000	64.30
1	90 745	59.06	95 300	66.46
5	86 491	57.90	94 038	63.33
10	84 762	54.03	93 365	58.77
20	81 638	45.88	91 760	49.70
30	76 619	38.57	88 923	41.13
40	71 897	30.77	85 958	32.37
50	65 702	23.17	81 123	23.97
60	56 548	16.06	71 870	16.35
70	41 680	9.85	54 111	9.92

* From United States Life Tables (pp. 208-209, 216-217), prepared by James W. Glover. Government Printing Office (*Washington*), 1921.

** From *Statistik Arsbok for Sverige* (p. 66), 1945.

Age span divided into three age groups. Dividing the age span into the three age groups 0-1, 1-5 and 5 years and over, a straightforward application of formulae given in (10) and (11) yields:

$$M_1 = \frac{1}{100{,}000}\,\frac{100{,}000}{100{,}000}[100{,}000(64.30-54.53) - 95{,}300(66.46-59.06)]$$
$$= 2.72$$

$$M_2 = \frac{1}{100{,}000}\,\frac{90{,}745}{95{,}300}[95{,}300(66.46-59.06) - 94{,}038(66.33-57.90)]$$
$$= 1.85$$

[4] I am grateful to Marianne Knopf of the Population Division, United Nations, for carrying out the calculations in this section.

and similarly:

$$M_3 = 4.70$$

$$M_{12} = 4.66$$

$$M_{23} = 6.72$$

$$M_{13} = \frac{1}{100{,}000}[100{,}000(64.30\text{-}54.53) - 95{,}300(66.46\text{-}59.06) + \frac{95{,}300}{90{,}745} \times 86{,}491(63.33\text{-}57.90)] = 7.65$$

$$M_{123} = 9.77$$

$$I_{12} = \frac{1}{100{,}000}\left(\frac{100{,}000}{100{,}000} - \frac{90{,}745}{95{,}300}\right)[95{,}300(66.46\text{-}59.06) - 94{,}038(63.33\text{-}57.90)] = 0.09$$

$$I_{23} = 0.17$$

$$I_{13} = \frac{1}{100{,}000} 86{,}491 \left(\frac{95{,}300}{90{,}705} - 1\right)(63.33\text{-}57.90) = 0.23$$

$$I_{123} = \frac{1}{100{,}000}(90{,}705 - 95{,}300)\left(\frac{86{,}491}{90{,}705} - \frac{94{,}038}{95{,}300}\right)(63.33\text{-}57.90) = 0.008$$

Application of the formulae given in equation (12) leads to

$$O_1 = \frac{1}{100{,}000}\,\frac{100{,}000}{100{,}000}\,[100{,}000(64.30\text{-}54.53) - 90{,}745(66.46\text{-}59.06)] = 3.05$$

Similarly:

$$O_2 = 2.12$$

$$O_3 = 5.11.$$

From the numerical values given above it is seen that

$$M_1 + I_{12} + I_{13} + I_{123} = 2.72 + 0.09 + 0.23 + 0.008 = 3.05 = O_1,$$

$$M_2 + I_{12} + I_{23} + I_{123} = 1.85 + 0.09 + 0.17 + 0.008 = 2.12 = O_2,$$

$$M_3 + I_{13} + I_{23} + I_{123} = 4.70 + 0.23 + 0.17 + 0.008 = 5.11 = O_3,$$

verifying the formulae given in equation (12).

Age span divided into a large number of groups. The formulae for the calculation of the "main effect," "operative effect," "effect-interaction deferred" and "effect-interaction forwarded" for any age group x to $x+n$ years as given in the text are as follows:

Main effect $= (1/l_0)\,(l_x/l'_x)\,[l'_x(e'_x - e_x) - l'_{x+n}(e'_{x+n} - e_{x+n})]$

Operative effect $= (1/l'_0)\,(l'_x/l_x)\,[l_x(e'_x - e_x) - l_{x+n}(e'_{x+n} - e_{x+n})]$

Effect-interaction deferred $= (1/l'_0)\,[l'_x(e'_x - e_x) - l'_{x+n}(e'_{x+n} - e_{x+n})]$

Effect-interaction forwarded $= (1/l_0)\,[l_x(e'_x - e_x) - l_{x+n}(e'_{x+n} - e_{x+n})]$.

Application of these formulae to the data given in Table 1, in which the age span is divided into ten age-groups, leads to the results given in Table 2 below.

TABLE 2

The various "effects"

Age group in years	Main effect	Operative effect	Effect-interaction deferred	Effect-interaction forwarded
0-1	2.72	3.05	2.72	3.05
1-5	1.85	2.12	1.95	2.02
5-10	0.63	0.74	0.68	0.68
10-20	0.84	0.99	0.92	0.90
20-30	1.09	1.30	1.23	1.16
30-40	0.78	0.94	0.90	0.81
40-50	0.61	0.75	0.73	0.62
50-60	0.36	0.45	0.44	0.36
60-70	0.13	0.17	0.17	0.13
70-	0.03	0.04	0.04	0.03
Total	9.04	10.55	9.78	9.76

The effects with interaction deferred or forwarded when summed over all the age groups equal the difference of the expectation of life at birth in the two periods, viz. 9.77, as they should. The sum of the "main effects" equals 9.04, the difference between this figure and 9.77,

viz. 0.73, being equal to the sum of the "interactions." As noted in the text the real difficulty is the lack of a logical basis for apportioning this interaction between the various age groups. The sum of the "operative effects" exceeds 9.77 by practically the same amount by which the sum of "main effects" is in defect. This feature is not algebraically correct but in practice will hold good due to the fact that the second and higher order interactions will in general be of negligible magnitude.

A BLOOD PRESSURE STUDY OF THE NATIVES OF PONAPE ISLAND, EASTERN CAROLINES

BY RUPERT I. MURRILL *

IN the summer of 1947, while making an anthropometric survey of the natives of Ponape in the Eastern Carolines, I had the opportunity at the same time of taking their blood pressures.

Ponape is a high basalt island approximately 900 miles southeast of Guam. The climate shows an annual average temperature of 80° F, humidity 86 per cent, and rain-fall around 180 inches — in short, a tropical island. Ponape is about 130-145 square miles in size, with mountains in the interior rising to 2500 feet. Between the coast line and the foot of the mountains the land is mostly flat, and it is here that the natives live. It is divided into 5 Districts,—Sokas, Net, U, Matolenim, and Kitti. Since the main town of Colonia was destroyed by incendiary bombs and has not been rebuilt, one can say that all the natives are essentially "rural."

SAMPLE

I particularly want to stress the nature of the sample. The natives may be conveniently divided into two main types: (1) "Pure" Ponapeans—that is to say, those whose ancestry shows no mixture with whites; and (2) Mixed. These can be subdivided into: (a) Ponapean natives who mixed with whites, such as Americans, English, French, Germans, and Portuguese; (b) Ponapean natives who mixed with out-islanders, natives from surrounding islands; (c) Ponapean natives who mixed with the Japanese. The present sample is concerned entirely with type (1). It is therefore representative of type (1) and not the total population of the island.

Actually one wonders whether the physiological phenomenon of blood pressure is any different in the above two types, as they are all living in the same environment.

* Research Fellow, University of Puerto Rico.

Since the natives live in separate farmsteads, and one has to walk to get around, distances are too great to visit each house individually. The local District Police Chief was therefore instructed to summon the men and women to the District meeting house, from about every third or fourth home. Great stress was laid on the fact that they were not to be from the same family, all this being carefully checked in the subsequent interview. This was done in each of the five Districts, so one can say that this was a random sample. Furthermore each one of the natives used was normal and healthy; that is to say, none of them were hospital patients nor was a single blood pressure reading taken in a hospital.

Blood pressure readings were taken on 127 men aged 20-59, and 124 women aged 18-59. This is approximately one-eighth of the "Pure" Ponapean population between the same ages.

METHOD

Before each subject was measured, he or she sat at a table being interviewed in regard to age and genealogy, for at least 15 minutes. They were put at ease as much as possible. At the end of this time the blood pressure was taken. The subject's arm rested on the table at about heart level. One reading was taken by the auscultatory method, always by myself, and always on the right arm, using a Tycos aneroid sphygmomanometer with a 12 cm. arm band. The systolic reading was taken at the appearance of the first sound and the diastolic at the change in tone.

RESULTS

In spite of the fact that it is claimed in recent papers (1) that blood pressure increases with age, especially after the age of 40, I have followed the example of Robinson and Brucer (2) and removed from the main sample those blood pressures of 140 systolic and over, and 90 diastolic or over. These pressures are provisionally designated as hypertensive — "abnormal," if you will. The remainder are called "normal" and the results are summarized in Tables 1 and 2. The complete data for all individuals are given in the Appendix Tables 1 A and 1 B.

By this method 10 males, or 7.9 per cent of the total male sample, and 14 females, or 11.3 per cent of the total female sample, have been excluded. Extremely low systolic pressures have not been discarded since none of the males and only 0.8 per cent of the females in the total group

had such pressures below 90 mm.; 92.1 per cent of all the males and 88.7 per cent of all the females had diastolic pressures below 90 mm.

TABLE 1

"Normal" systolic and diastolic blood pressures by age—males (delimited group)

mm. Hg	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	ALL AGES
					Systolic pressure				
90-99		1			4	2	1	2	10
100-109	5	5	11	3	6	4	3	3	40
110-119	5	13	9	4	4	6	4	3	48
120-129	1	2	3	2	3	2	2	2	17
130-139						2			2
Total	11	21	23	9	17	16	10	10	117
Means	110.4	111.6	110.5	112.9	107.5	112.7	111.0	109.0	110.7
					Total group				
Means	110.4	112.2	112.8	115.0	109.8	112.7	112.3	109.5	111.9
					Diastolic pressure				
50-59	2			1				1	4
60-69	2	1	5	1	5	2		2	18
70-79	4	14	11	4	8	7	8	4	60
80-89	3	6	7	3	4	7	2	3	35
Total	11	21	23	9	17	16	10	10	117
Means	71.3	76.4	74.9	74.0	73.4	77.1	79.8	73.0	74.8
					Total group				
Means	71.3	77.2	77.1	77.0	75.6	77.1	79.8	74.9	76.4

Here and elsewhere groups have been centered at 94, 104, etc., for the calculation of means.

In the delimited group (Table 1), for 117 males the mean systolic pressure is 110.7 mm. and the standard deviation 7.8 mm. Therefore two-thirds of the men have a systolic pressure between 102.9 and 118.5 mm. The mean diastolic pressure is 74.8 mm. with a standard deviation of 5.8 mm., giving a range of 69.0 to 80.6 mm.

For 110 females in the delimited group (Table 2), the mean systolic pressure is 108.6 mm. with a standard deviation of 7.4 mm. and a range of 101.2 to 116.0 mm. The mean diastolic pressure is 74.1 mm., with a standard deviation of 5.1 mm., thus giving a range of 69.0 to 79.2 mm.

In the delimited group 83.7 per cent of the males and 87.3 per cent of the females have a systolic pressure below 120 mm.

TABLE 2

"Normal" systolic and diastolic blood pressures by age—females (delimited group)

mm. Hg	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-59	ALL AGES
				Systolic pressure					
80-89			1						1
90-99	1		3	3		2		1	10
100-109	2	9	10	9	7	5	4	4	50
110-119	1		7	8	6	5	5	3	35
120-129	1	1	2		3	3	2	2	14
Total	5	10	23	20	16	15	11	10	110
Means	108.0	106.0	106.6	106.5	111.5	110.0	112.2	110.0	108.6
				Total group					
Means	108.0	107.6	107.3	107.3	111.7	110.9	117.1	119.0	111.0
				Diastolic pressure					
50-59			1		1				2
60-69		3	6	5	3	4	1		22
70-79	5	5	9	14	9	8	4	5	59
80-89		2	7	1	3	3	6	5	27
Total	5	10	23	20	16	15	11	10	110
Means	74.0	73.0	73.6	72.0	72.7	73.3	78.5	79.0	74.1
				Total group					
Means	74.0	74.9	74.4	73.0	74.0	74.6	85.3	83.5	76.6

For those who will have none of these delimited groups, Table 3 has been added. In this table none of the pressures are excluded, yet the mean systolic pressure for males is 111.9 mm., with a mean diastolic pressure of 76.4 mm. In the females the mean systolic pressure is 111.0 mm., and the mean diastolic 76.6 mm. There is little difference then between this group and the delimited group.

The effect of age on blood pressure in the delimited group is shown in Fig. 1.

TABLE 3

Systolic and diastolic blood pressures of total group by sex

Systolic pressure			Diastolic pressure		
mm.	Males	Females	mm.	Males	Females
80-89	–	1	50-59	4	2
90-99	10	10	60-69	18	22
100-109	40	50	70-79	60	59
110-119	50	37	80-89	35	27
120-129	22	21	90-99	8	11
130-139	4	2	100-109	2	3
140-149	1	2			
150-159	–	–			
160-169	–	1			
Total	127	124	Total	127	124
Mean	111.9	111.0	Mean	76.4	76.6
% under 100	7.9	8.9	% under 60	3.1	1.6
% under 110	39.4	49.2	% under 70	17.3	19.4
% under 120	78.7	79.0	% under 80	64.6	66.9

This graph indicates that there are differences in the pressures of males and females in this sample. Up to 40 years of age, the females have lower systolic and, on the whole, lower diastolic pressures than the males.

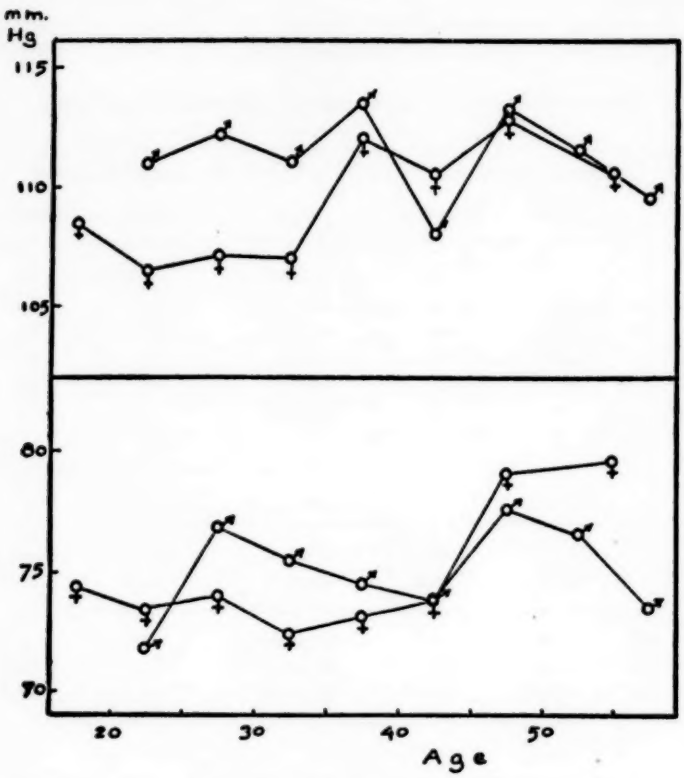

FIG. 1. BLOOD PRESSURE AND AGE, BY SEX, DELIMITED GROUP. UPPER PANEL, SYSTOLIC; LOWER PANEL, DIASTOLIC

After the age of 40, the systolic pressures of the males and females are similar while the diastolic pressures are somewhat higher in the females. Table 4 shows that more males over 40 years of age have systolic pressures under 100 and 110 mm., while the reverse is true in the females.

TABLE 4

Systolic blood pressure in males and females (delimited group)

Systolic pressure (mm. Hg)	Males		Females	
	Under 40	Over 40	Under 40	Over 40
Under 100	1.6%	17.0%	10.8%	8.3%
Under 110	39.1	47.2	60.8	44.4
Under 120	87.5	79.2	90.5	80.6
Over 120	12.5	20.8	9.5	19.4

Table 5, which includes all pressures in those over 40 years of age, shows that as far as the mean systolic pressure is concerned, in the males it increases up to 50 years of age then decreases. In the females it increases with age throughout the age groups. The diastolic pressure

TABLE 5

Systolic and diastolic blood pressures by sex and age, of those 40 years of age and over (total group)

Male			Female		
Age	No.	Mean	Age	No.	Mean
		Systolic			
40-44	19	110.3	40-44	16	111.4
45-49	16	113.2	45-49	16	117.6
50-54	12	112.8	50-59	14	119.5
55-59	11	110.0			
		Diastolic			
40-44	19	76.1	40-44	16	75.1
45-49	16	77.6	45-49	16	85.8
50-54	12	80.3	50-59	14	83.8
55-59	11	75.4			

in the males increases up to 55 years of age, then decreases; in the females it increases up to 50 years of age, then decreases. Therefore in this sample, except for the mean systolic pressure in the females (Table 5), and the female mean diastolic pressure after the age of 30

in the delimited group (Fig. 1), there is no clear-cut increase of systolic or diastolic pressure with age.[1]

Finally it is notable that in those males and females excluded from the delimited groups, the majority were excluded because they had a high diastolic pressure rather than a high systolic pressure. This seems to agree with Huber (3) and Russek (1) that the diastolic level is more indicative of hypertension.

DISCUSSION

This paper is concerned not with hypertension but rather with so-called hypotension. Several authors (4) have indicated that blood pressure in the tropics is "low" compared to American "normal" pressure. These "low" systolic pressures found in the tropics range anywhere from 100 to 125 mm. for males or females. If, however, we use Robinson and Brucer's "normal" range (at least for those up to 40 years of age) of 100-125 mm. systolic for males or females, even though their sample is by no means representative of the population of the United States, where, then, is the "low" pressure ascribed to the tropics? Why any need to introduce such explanatory factors as race, diet, climate, and so on?

However, as can be seen in this sample, 42.7 per cent of the males and 55.5 per cent of the females in the delimited group have systolic pressures under 110 mm., and it may be argued that regardless of the "normal" range, this is a high proportion of "low" pressures.

It is admitted here that this study has certain limitations. First, only one blood pressure reading was taken per person. Second, it is estimated that 100 per cent of the Ponape population is infested with hookworms, in addition to which a large proportion are suffering from tuberculosis (malaria is not present on Ponape). In regard to this second point, one might say, therefore, that this sample is neither "normal" nor "healthy." One immediately wonders what effect this has on the

[1] [NOTE. It has seemed of some interest to examine the statistical significance of the changes of blood pressure with age. Analysis of variance, using the total group, gives results summarized below.

For the males, there is no evidence of any significant variation with age in either systolic or diastolic pressure.

For the females there are significant changes with age in both systolic and diastolic pressure. The nature of these changes consists in an increase in both systolic and diastolic pressure at ages 45 and over. These changes are not apparent in the delimited group. Actually, 9 of the excluded cases occur in the two highest age groups.—C. P. W.]

blood pressure. It is possible that this might explain the proportion of the sample with pressures below 110 mm. systolic.

Donnison (4) reports mean systolic pressures in male Negroes in Kenya, East Africa of 123.1 mm. at 15-19 years of age, increasing to 126.1 mm. at 30-34 and thereafter decreasing to 105.8 mm. at 60 years of age and over. Here we have not only blood pressure decreasing with age, but also an area with a high incidence of schistosomiasis. One wonders again what could be the effect of such infestation on blood pressure. Saunders (4) reports 80 per cent infestation of the population with ascariasis in Yucatan. The time has come to cease wondering. A study on blood pressure should be done, in an area having intestinal parasites, using one group that is known to be free of the parasites as control, and another group that does have them.

It is claimed that when foreigners migrate to tropical countries their blood pressure drops considerably. Leaving aside their new level of blood pressure and its comparison with American "norms," one can find a simpler explanation for this lowering of blood pressure, naïve as it may sound. Most foreigners in tropical or sub-tropical countries live well compared to their former life; they usually have servants, they hurry less,—in short, they slow down. This lessened tension may well lower the blood pressure.

It is interesting to observe that several authors (5) claim a high incidence of hypertension among Negroes both in the United States and the tropics, contrary to Donnison's results. In each case we have an area where the Negro is in close contact with white civilization. He may be living economically and emotionally on an inferior plane to the whites and this may well affect his blood pressure. Saunders (5) states, "A greater percentage of Negroes in the Virgin Islands have a systolic pressure of more than 150 mm., and of less than 110 mm., than residents of the United States." He seems to have two groups within the Negro population. It would be worth knowing which Negroes (urban or rural, etc.) in his sample had the high pressures.

In conclusion, and this is a main point of this paper, it should be evident by now that in all the voluminous literature on blood pressure, not one sample exists that is truly representative of the normal healthy population of the United States. And yet one finds published samples, half of which contain hospital patients, or samples almost wholly consisting of urban dwellers, or insurance company samples, etc., being passed off as "representative" of the population of the United States—not to mention the rather dogmatic conclusions made from such samples

concerning blood pressure. Mere numbers prove nothing. It must be obvious that blood pressures of a large number of Belgian Congo pygmies are not representative of blood pressures of all African Negroes, yet this is, in essence, the technique used in the medical literature.

The medical profession is greatly concerned with diseases of the heart and its concomitant results. We now have a phrase for it, "The No. 1 Killer." Surely it is high time that blood pressures be taken on a sample that *is* representative of the population of the United States (in regard to sex, age, occupation, urban or rural residence, diet, and so on), and settle once and for all what is "normal" blood pressure, and whether or not blood pressure does increase with age. Such an undertaking will be expensive and time-consuming but certainly worthwhile.

SUMMARY

1. A random representative sample of "Pure" Ponapeans has been used in this study.

2. The mean systolic blood pressure for 127 males, ranging in years from 20 to 59, was found to be 111.9 mm., the diastolic pressure 76.4 mm. For 124 females, ranging in age from 18 to 59 years, the mean systolic pressure was 111.0 mm., the diastolic 76.6 mm.

3. In a delimited group (excluding pressures of 140 mm. systolic or over and 90 mm. diastolic or over) the mean systolic pressure for 117 males was 110.7 mm., the diastolic 74.8 mm. For 110 females, the mean systolic pressure was 108.6 mm., the diastolic 74.1 mm.

4. Approximately 80 per cent in the total group, and over 80 per cent in the delimited group, of males and females, had systolic pressures below 120 mm.

5. Depending on what range one uses for "normal" pressure, it is doubted if tropical hypotension exists.

6. It is hoped that truly "representative" samples will be gathered in the future.

BIBLIOGRAPHY

1. (a) MASTER, A. M., *et al.* *Journ. Am. Med. Assn.*, 1943, 121, 1251.
(b) RUSSEK, H. I., *et al.* *Am. Heart Journ.*, 1946, 32, 468.
2. (a) ROBINSON, S. C., and BRUCER, M. *Arch. Int. Med.*, 1939, 64, 409.
3. (a) HUBER, E. G. *Human Biology*, 1933, 5, 541.
(b) RUSSEK, H. I., *et al.* *Vide supra.*

4. (a) ASHFORD, B. K., and DOWLING, G. B. *Puerto Rico Journ. Pub. Health and Trop. Med.*, 1930, 5, 477.
(b) CONCEPCION, I. and BULATAO, E. *Philippine Journ. Sci.*, 1916, 11, 135.
(c) CADBURY, W. W. *Arch. Int. Med.*, 1922, 30, 362.
(d) DONNISON, C. P. *Lancet*, 1929, 1, 6.
(e) FOSTER, J. H. *Arch. Int. Med.*, 1927, 40, 38.
(f) SAUNDERS, G. M. *Am. Journ. Med. Sci.*, 1933, 185, 843.
5. (a) ADAMS, J. M. *Am. Journ. Med. Sci.*, 1932, 184, 342.
(b) FLAXMAN, N. *Am. Journ. Med. Sci.*, 1934, 188, 639.
(c) KEAN, H. *Arch. Int. Med.*, 1941, 68, 466.
(d) SAUNDERS, G. M. and BANCROFT, H. *Am. Heart Journ.*, 1942, 23, 410.

APPENDIX

TABLE 1A

Age and blood pressures (in mm.) of 127 male natives of Ponape Island
(Individuals marked * were excluded from the delimited group treated in the text.)

AGE	SYSTOLIC	DIASTOLIC	AGE	SYSTOLIC	DIASTOLIC
20	100	58	38	122	70
	110	72	39	104	82
	112	50		106	70
	118	78	40	92	68
	120	80		104	68
21	108	72		114	78
	108	80		114	82
22	110	62	41	108	78
	110	70		*124	98
24	104	68	42	98	70
	108	80		104	72
25	118	72		106	70
26	104	78		108	70
	108	72		120	80
	110	70		*130	92
	110	78	43	94	74
	120	76		96	68
27	104	78		100	60
	112	78		122	86
	114	78		126	84
	114	80	44	112	68
	116	76		114	78
	118	80	45	104	74
	118	86	46	102	70
	*122	90		110	78
28	98	70	47	98	72
	102	68		102	74
	106	72		110	80
	110	78		118	82
	112	80		136	82
	122	84	48	98	68
29	118	70		102	72
	118	80		116	82
30	104	66		116	82
	110	80		120	82
	110	84		130	70
	118	78	49	118	64
	118	80		120	80
31	100	62	51	*112	90
	106	70	52	102	78
	110	78		104	70
	120	80		124	88
32	106	66		*124	100
	114	78	53	110	78
	116	76		118	78
	120	84		120	85
33	104	70	54	96	70
	108	68		102	70
	108	70		110	72
	108	78		110	78
	114	80	55	104	68
	116	80		116	84
	122	84		122	82
	*122	90	56	100	76
	*140	90		*112	96
34	104	74	57	104	64
	106	76		118	80
	108	68	58	96	58
	*122	92		98	70
35	110	76		110	70
	120	82		124	78
37	104	68			
	110	78			
	110	88			
	114	58			
	*132	100			

TABLE 1B

Age and blood pressures of 124 female natives of Ponape Island

(Individuals marked * were excluded from the delimited group treated in the text.)

AGE	SYSTOLIC	DIASTOLIC
18	98	72
	106	72
	118	78
	124	78
19	102	70
20	104	78
	124	88
21	100	70
	*124	90
22	102	68
	102	68
	102	70
	108	80
23	104	74
	106	72
24	100	64
25	84	64
	98	56
	100	70
	110	78
	114	80
	118	78
	118	80
	122	80
	122	80
26	108	84
	114	80
27	98	60
	100	72
	104	60
	*122	92
28	100	76
	104	62
	104	76
	106	68
	110	70
	110	76
29	90	64
	102	70
	104	82
30	102	64
	104	76
	108	72
	108	72
	108	74
	108	78
	114	74
	116	80
	118	64
	*122	98
31	98	64
	102	68
32	110	70
	110	72
	112	78
	118	76
33	98	72
	100	68
	104	78
34	94	76
	110	78
35	102	58
	102	78
	104	76
	110	74
	120	78
36	102	76
	114	64
	114	72
	*118	92
37	102	76
	110	78
	120	80
38	100	62
	110	76
	110	80
39	102	62
	124	82
40	100	64
	110	72
	110	82
	112	74
41	98	60
	120	78
	122	84
42	100	72
	108	72
43	98	66
	100	62
	112	76
	114	78
	120	80
44	102	78
	*122	90
45	100	70
	102	70
	106	80
	*138	102
46	108	68
	118	82
	*120	100
	124	76
47	*120	96
	124	88
	*144	102
48	110	78
	110	80
	114	84
49	*116	92
	118	80
50	98	76
	106	80
	118	84
	128	88
	*130	90
	*160	90
51	106	74
52	102	72
	110	72
	118	80
	*126	90
53	106	76
56	124	82
	*144	90

NOTES

REPORT ON A SERIES OF ILLEGAL ABORTIONS INDUCED BY PHYSICIANS*

SEVERAL investigators (1, 2, 4, 5, 7) have presented evidence to show that at least in urban areas of the United States the great majority of illegal abortions are induced by physicians. These cases are rarely admitted to hospitals for completion or for the treatment of complications, and they do not often appear in studies of mortality. Little is known about them except that they are very numerous.

Through good fortune the author has come into possession of a series of 363 records representing the practice of two such abortion specialists in a large Eastern city during several months in 1948. Thanks are due to the individuals who have made this material accessible for research. The reader must be satisfied with the author's assurance that the records are genuine, complete, and probably as good as those kept by most practitioners and clinics. However, the material does not lend itself to a study of the course of recovery or of sequelae.

All but one of these abortions were performed between the 5th and 13th weeks of pregnancy, counting from the estimated date of conception. The average period of gestation was 8.5 weeks. One pregnancy was interrupted in the 15th week. It is not known why this particular patient was accepted at that advanced stage.

The records do not contain usable information concerning the occupational, social, or economic background of the clientele. It is known, however, that the fees charged varied from $300 to $500, which should indicate a fairly high level of income. Of the 363 women, 102 were single, 180 (49.6 ± 2.6 per cent) married, and 81 previously married. About half of the latter group were widows, the others divorced. The percentage of married women was lower than, but not significantly different from that reported by Hamilton (3) in New York, and much lower than

* Read at the annual meeting of the American Population Association, Princeton, N. J., May 28, 1949.

among the cases studied by Simons (6) in Minneapolis, who found three-fourths of the induced cases in the married group. Since his were hospital patients their abortions had been induced either by the pregnant woman herself or by a midwife or other non-medical abortionist. Many were probably drawn from the underprivileged group where economic pressure on the married woman is comparatively greater, and social pressure on the unmarried woman perhaps less, than in the class from which the present series is recruited. Another factor contributing to the difference may be a greater willingness on the part of married women to seek admission to a hospital for the treatment of abortion, whereas the unmarried will tend to avoid hospitalization unless it is clearly necessary because of fever or other complications.

TABLE 1

Women undergoing illegal abortion by marital status, age, number of children ever born, and previous induced abortions

	Single	Married	Previously married	Total
Age (years)				
10-14	3	–	–	3
15-19	22	10	2	34
20-24	29	44	15	88
25-29	21	42	28	91
30-34	18	42	20	80
35-39	6	19	10	35
40 and over	3	23	6	32
Children born				
0	100	45	25	170
1	2	39	31	72
2	–	52	16	68
3	–	27	8	35
4 and over	–	17	1	18
Previous abortions				
0	87	147	64	298
1	10	18	15	43
2	2	10	2	14
3 and over	3	5	–	8
Total	102	180	81	363

The women in the present series ranged in age from 13 to 47 years (Table 1). The average age was 28.6 years. As one would expect, the single women were youngest; while for the married and the previously married, the average age was practically the same.

Of the group as a whole, 170 or almost half had never borne a child (Table 1). The childless included all but two of the single women, one-fourth of the married, and one-third of the previously married. The total number of children born was 397 for the whole group, including 303 reported by the married women. The average number of children ever born was 1.09 for the group as a whole and 1.68 for the married.

It is of interest to compare the married women in this series with a group of women delivered on the private service of a large hospital in the same city in 1948. Table 2 presents a comparison in terms of children born before the current abortion or delivery. The differences are

TABLE 2

Comparison of 180 married women undergoing illegal abortion and 481 women delivered on a private service in 1948

Previous children	Number of cases		Percentage distribution		Ratio of percentages
	Abortions	Deliveries	Abortions	Deliveries	
0	45	206	25.0	42.8	3 : 5
1	39	164	21.7	34.1	2 : 3
2	52	86	28.9	17.9	5 : 3
3	27	21	15.0	4.4	3 : 1
4 and over	17	4	9.4	.8	11 : 1
Total	180	481	100.0	100.0	–

striking. They confirm earlier findings (1, 4, 8, 9) that the incidence of illegal induced abortion tends to increase with parity, and suggest that in this group abortion was used primarily not as a method of child-spacing but of limiting the ultimate size of the family. This was probably not true for the ten married women, 15-19 years old, in the abortion series. There was only one case in this age group among the 481 deliveries. This seems to indicate that illegal abortions are also resorted to by those who are not yet ready to shoulder the responsibilities of raising a family.

In addition to their 397 children, the women in this series reported 29 unintentional abortions. This number represents seven per cent of all

pregnancies terminating without interference. This is a plausible ratio which is only slightly lower than one would expect on the basis of earlier studies (5, 10). It adds to the general credibility of the records.

Sixty-five women, or 17.9 per cent of the group, admitted one or more induced abortions preceding the current one (Table 1). The proportion was not significantly different for the single, the married, and the previously married. This information is probably more complete than it is likely to be in a series of different origin because there was apparently less reason for dissimulation. The total number of previous induced abortions was 101, representing one-fifth of all previous pregnancies. Many of them had been performed by the same operators as the current one.

If the proposition is accepted that first and later induced abortions had an equal chance of being represented in our sample, then the average number of induced abortions in the population from which this sample is drawn can be estimated to be on the order of 1.2 (363/298) per woman. This figure, of course, refers only to those women who had at least one induced abortion during their reproductive lives. Since many women never resort to illegal abortion, the average number for all women must have been much lower, but it cannot be computed from this series of records.

As far as the author is aware, this is the first series of illegal abortions induced by physicians ever reported on in American medical literature. The number of cases is not large and it cannot be estimated how far the findings are typical of the activities of abortion specialists around the country. Much work remains to be done until we shall have adequate knowledge of this much neglected field of human behavior.

REFERENCES

1. Brunner, E. K. The outcome of 1556 conceptions. *Human Biology*, 13(2): 159-176, 1941.
2. Brunner, E. K., and Newton, L. Abortions in relation to viable births in 10,609 pregnancies. *Am. J. Obst. and Gyn.*, 38(1): 82-90, 1939.
3. Hamilton, V. C. Some sociologic and psychologic observations on abortion. *Am. J. Obst. and Gyn.*, 39(6): 919-928, 1940.
4. Kopp, M. Birth Control in Practice. *New York*, 1934.
5. National Committee on Maternal Health. The Abortion Problem. *Baltimore*, 1944.
6. Simons, J. H. Statistical analysis of one thousand abortions. *Am. J. Obst. and Gyn.*, 37(5): 840-849, 1939.

7. STIX, R. K. A study of pregnancy wastage. *Milbank Memorial Fund Quarterly*, 13(4) : 347-365, 1935.
8. STIX, R. K., and NOTESTEIN, F. W. Controlled Fertility. *Baltimore*, 1940.
9. WHELPTON, P. K., and KISER, C. V. The comparative influence on fertility of contraception and impairments of fecundity. *Milbank Memorial Fund Quarterly*, 26(2) : 303-357, 1948.
10. WIEHL, D. G. A summary of data on reported incidence of abortion. *Milbank Memorial Fund Quarterly*, 16(1) : 1-9, 1938.

CHRISTOPHER TIETZE, M. D.

615 North Wolfe Street
Baltimore 5, Maryland